Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013-2023)

Computers in Biology and Medicine 165 (2023) 107441

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023)

Silvia Seoni [a], Vicnesh Jahmunah [b], Massimo Salvi [a], Prabal Datta Barua [c,d], Filippo Molinari [a,*], U. Rajendra Acharya [e]

[a] Biolab, PolitoBIOMedLab, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
[b] School of Engineering (SEG), Nanyang Polytechnic, Singapore
[c] School of Business (Information System), University of Southern Queensland, Toowoomba, QLD, 4350, Australia
[d] Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, 2007, Australia
[e] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia

ARTICLE INFO

Keywords:
Uncertainty techniques
Machine learning models
Deep learning models
PRISMA
Images
Signals
Healthcare
Bayesian models

ABSTRACT

Uncertainty estimation in healthcare involves quantifying and understanding the inherent uncertainty or variability associated with medical predictions, diagnoses, and treatment outcomes. In this era of Artificial Intelligence (AI) models, uncertainty estimation becomes vital to ensure safe decision-making in the medical field. Therefore, this review focuses on the application of uncertainty techniques to machine and deep learning models in healthcare.

A systematic literature review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Our analysis revealed that Bayesian methods were the predominant technique for uncertainty quantification in machine learning models, with Fuzzy systems being the second most used approach. Regarding deep learning models, Bayesian methods emerged as the most prevalent approach, finding application in nearly all aspects of medical imaging.

Most of the studies reported in this paper focused on medical images, highlighting the prevalent application of uncertainty quantification techniques using deep learning models compared to machine learning models. Interestingly, we observed a scarcity of studies applying uncertainty quantification to physiological signals. Thus, future research on uncertainty quantification should prioritize investigating the application of these techniques to physiological signals.

Overall, our review highlights the significance of integrating uncertainty techniques in healthcare applications of machine learning and deep learning models. This can provide valuable insights and practical solutions to manage uncertainty in real-world medical data, ultimately improving the accuracy and reliability of medical diagnoses and treatment recommendations.

1. Introduction

Artificial intelligence (AI) has emerged as a promising technology with significant potential to transform the healthcare industry. AI technologies such as machine learning, natural language processing, and computer vision can analyze vast amounts of patient data and provide valuable insights to healthcare professionals. The use of AI in healthcare has the potential to revolutionize the way in which healthcare is delivered, improving patient outcomes, reducing costs, and increasing access to care. AI can assist healthcare providers in making more accurate diagnoses, predicting outcomes, and developing personalized treatment plans for patients. Additionally, AI-powered tools can help healthcare providers identify early warning signs of diseases and conditions, enabling early intervention and prevention [1]. This can greatly enhance the efficiency and effectiveness of healthcare delivery, ultimately leading to better health outcomes for patients.

Despite the promising potential of AI in healthcare, there are also concerns regarding privacy, security, and ethical considerations. As such, it is important to carefully consider the benefits and risks associated with the use of AI in healthcare and to ensure that the technology is

* Corresponding author.
E-mail address: filippo.molinari@polito.it (F. Molinari).

https://doi.org/10.1016/j.compbiomed.2023.107441
Received 4 August 2023; Received in revised form 27 August 2023; Accepted 29 August 2023
Available online 1 September 2023

deployed ethically and responsibly. Indeed, the 'black box' nature of these AI systems has raised concerns about their reliability and accountability [2]. The inner workings of these models are often not comprehensible to end-users, and even data scientists may struggle to interpret the algorithm [2]. This entire scenario makes it challenging for end-users to trust the AI system they are interacting with, potentially leading to skepticism or even rejection [2].

In response to this need for transparency and trust, the emerging field of explainable AI (XAI) employs techniques to enhance the interpretability of AI models [2]. XAI techniques are effective in uncovering the 'black box' aspect of machine learning models and providing explanations for the decisions they make [2]. However, while these techniques can improve the interpretability of AI models, they do not address the practical assessment of decision reliability [3]. Furthermore, XAI techniques do not capture the AI models' overconfident predictions and vulnerability to adversarial attacks [4], which can lead to user uncertainty about AI system prediction.

To ensure safety and reliability [5], it is crucial to evaluate the uncertainty of AI system predictions. The concept of uncertainty pertains to the level of confidence or ambiguity in the predictions generated by these models and can result from a variety of factors such as incomplete or noisy data, limited domain knowledge, or inherent randomness in the system, making it a crucial consideration in ensuring the reliability, interpretability, and safety of AI models. Providing uncertainty estimates in AI systems is essential for ensuring safe decision-making in high-risk domains characterized by diverse data sources, as seen in remote sensing [6]. Uncertainty estimates are also critical in domains where the nature of uncertainty is an essential part of the training methods, such as in active learning [7] and reinforcement learning [8]. By incorporating strategies for quantifying and communicating uncertainty in AI systems, we can enhance their effectiveness and foster greater trust in their predictions.

Predictive uncertainty is a widely used technique concerned with the uncertainty associated with making predictions or estimates using a model. It quantifies the level of confidence or reliability in the model's predictions for new or unseen data points. The most common approach for estimating predictive uncertainty involves modeling the uncertainty caused by the model itself (model or epistemic uncertainty) separately from the uncertainty caused by the data (data or aleatoric uncertainty) [9]. Aleatoric uncertainty is an intrinsic property of the data distribution [4] and arises in situations with a large amount of data that are not informative [10] or incomplete, noisy, conflicting, or multi-modal [11]. On the other hand, epistemic uncertainty occurs due to insufficient knowledge, a poor representation of the training data, or flaws in the model itself, leading to uncertainty about the model's behavior or performance in new or unseen situations [4]. While model uncertainty can be reduced by improving the architecture, learning process, and training data quality, data uncertainties are irreducible [4].

Predictive uncertainty can be classified into three main groups based on predictive uncertainty: *in-domain uncertainty* [12], *domain-shift uncertainty* [13], and *out-of-domain uncertainty* [14,15]. *In-domain uncertainty* refers to input data that is assumed to be drawn from the same distribution as the training data. This type of uncertainty arises when the model is unable to accurately predict an in-domain sample due to a lack of relevant knowledge. Additionally, design inaccuracies in the model can also contribute to in-domain uncertainty [12]. *Domain-shift uncertainty* [13] describes the uncertainty associated with input data that is drawn from a distribution that is shifted from the training distribution. This shift can be caused by poor representation of training data changes in real-world circumstances [13]. This shift may increase uncertainty because the deep model may struggle to explain the domain-shifted data based on the seen data used for training. *Out-of-domain uncertainty* [14,15] refers to the uncertainty associated with an input extracted from the subgroup of unknown data, wherein the distribution of unknown data is dissimilar and far from the distribution of training data. This type of uncertainty arises when the deep model is unable to explain an out-of-domain sample due to its lack of knowledge of the out-of-domain data [4].

As a result, model uncertainty encompasses what the deep model does not know due to the lack of in-domain or out-of-domain knowledge. This includes in-domain, domain-shift, and out-of-domain uncertainties. In contrast, data uncertainty only includes in-domain uncertainty caused by the nature of the data used to train the model [4]. Uncertainty can be introduced in healthcare in various ways (Fig. 1), for example:

- Variability in measurements: Medical measurements such as blood pressure, heart rate, and oxygen saturation can vary due to various factors such as measurement noise, biological variability, and measurement error.
- Incomplete or missing data: Medical data collected from patients may be incomplete or missing due to various reasons such as incomplete medical records, data entry errors, or patient non-compliance.
- Uncertainty in medical diagnosis: Medical diagnosis involves making decisions based on incomplete information and subjective interpretation of medical data, which may introduce uncertainty in the diagnosis.
- Uncertainty in medical treatment: Medical treatment involves making decisions based on uncertain outcomes and potential side effects, which may introduce uncertainty in the treatment process.

This review paper provides a comprehensive overview of uncertainty estimation in healthcare. The paper reviews recent advances in the field, highlights current challenges, and identifies potential research opportunities. In addition to providing a general outline of uncertainty quantification methods applied in the machine and deep learning models, the paper also discusses the most prevalent, emerging, and technically promising techniques in this research field.

2. Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were closely followed to select the most relevant articles on uncertainty estimation methods applied to healthcare, using traditional machine learning and advanced deep learning models.

2.1. Related reviews

The topic of uncertainty is highly relevant in the field of data analysis, and several reviews have recently been published on the subject. However, these reviews have limitations in terms of their scope and focus:

- Broekhuizen et al. [16] "A Review and Classification of Approaches for Dealing with Uncertainty in Multi-Criteria Decision Analysis for Healthcare Decisions". In this review, the authors discuss techniques for estimating uncertainty in multi-criteria decision analysis for healthcare decisions, without focusing on machine and deep learning approaches or considering only medical data.
- Lambert et al. [17] "Trustworthy clinical AI solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis". Here the authors focus only on deep learning approaches, neglecting machine learning ones. Moreover, the focus is only on medical images.
- Loftus et al. [18] "Uncertainty-aware deep learning in healthcare: A scoping review". The authors evaluate methods for quantifying uncertainty in deep learning for healthcare applications but analyze relatively few studies (around 30 papers).
- Gawlikowski et al. [19] "A Survey of Uncertainty in Deep Neural Networks". The authors provide a comprehensive review focusing

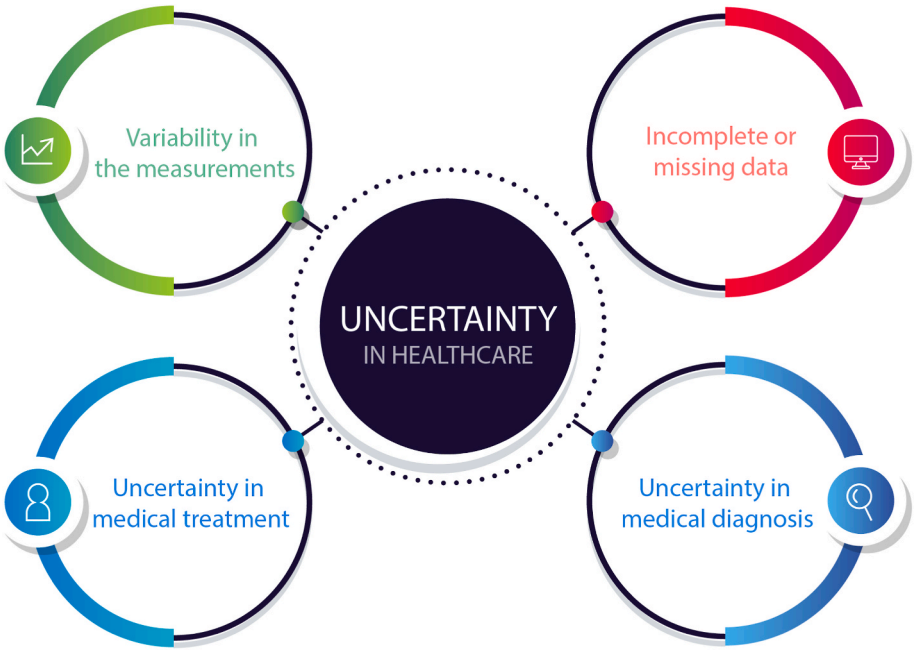

Fig. 1. Different sources of uncertainty possibilities in healthcare such as variability in measurements, incomplete or missing data, uncertainty in medical diagnosis, and uncertainty in medical treatment.

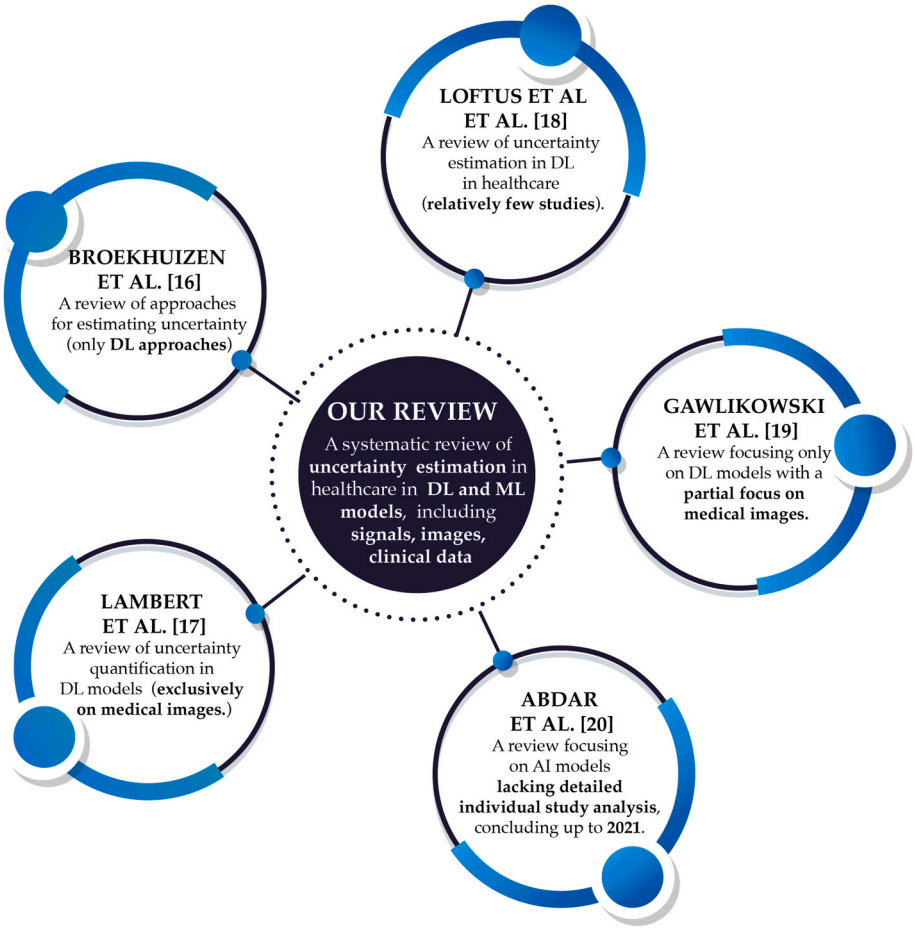

Fig. 2. Comparison between our review paper and the existing literature reviews.
**ML and DL are machine learning and deep learning.

only on deep neural networks, with a partial focus on medical images.
- Authors in Ref. [20] investigated UQ techniques in AI models and provided overview without delving into individual study nuances or explicitly distinguishing between ML and DL methods.

Thus, there is a need for a comprehensive review that covers both machine learning and deep learning approaches and analyzes all types of medical data, including physiological signals and medical images. This review aims to provide an overview of uncertainty quantification techniques applied in healthcare, with a focus on both machine learning and deep learning frameworks. Fig. 2 shows how our review integrates the existing literature reviews, providing an overview of all the works on uncertainty estimation in healthcare.

2.2. Search strategy

Only articles published in the last decade (2013–2023) were included in this review. The appropriate journal articles were searched through the Institute of Electrical and Electronics Engineers (IEEE), Google Scholar, PubMed, and Scopus scientific repositories. For the retrieval of articles focusing on machine learning and deep learning, the Boolean search strings such as "Uncertainty estimation", "Human healthcare", "Signals", "Images", "Machine learning" and "Deep learning" were used in various combinations from Scopus and PubMed scientific repositories. Two distinct searches were performed: one focusing on uncertainty estimation methods based on machine learning, and the other on those based on deep learning. The search was conducted between September 2022 to January 2023.

2.3. Study selection and quality assessment

A total of 424 articles and 553 articles were identified, respectively, using Boolean search strings for machine learning-based methods and deep learning-based methods. About 74 (ML) and 96 (DL) duplicate and irrelevant articles on deep learning were eliminated wherein articles on 'animal health' or 'model explainability'. Theses, books, and abstracts were also excluded. Thereafter, studies were included if they met the following criteria:

(i) They described uncertainty estimation methods used in healthcare involving human data (images/signals),
(ii) They described uncertainty estimation methods used in healthcare, based on machine learning or deep learning models,
(iii) They were published between the years 2013 and 2023,
(iv) They were published in a peer-reviewed journal,
(v) They were published in English.

Articles were excluded if they were: (i) not written in English, (ii) a review article or pilot study, (iii) an abstract or a book chapter, (iv) too similar to other studies, (v) published before 2013, or (vi) not available in full text. After careful examination, 312 articles for machine learning and 350 articles for deep learning were excluded based on the aforementioned criteria. The final selection yielded 38 articles for machine learning and 107 articles for deep learning, focusing on uncertainty estimation methods in healthcare. Table 1 provides a summary of these articles, and Fig. 3 depicts the utilization of the PRISMA guideline in article selection for this review.

Table 1
Results of the Boolean search string for the respective repositories on uncertainty estimation methods using machine learning.

Uncertainty techniques	Boolean search string			
	Database	Title	AND [Title/ Abstract/Full text]	No. of articles
Machine learning-based	IEEE, Scopus, PubMed	"Uncertainty estimation for human healthcare "	Machine learning model/signals/ images	38
Deep learning-based	IEEE, Google Scholar, PubMed	"Uncertainty estimation for human healthcare "	Deep learning model/signals/ images/imaging applications/non-imaging applications	107

3. Results

3.1. Uncertainty quantification in machine learning

Effective management of uncertainty is a crucial factor in medical decision-making, particularly in the context of diagnostic procedures. Table 2 illustrates the distribution of works based on the employed method for uncertainty management in machine learning approaches. According to the research papers, it can be concluded that the most utilized algorithms for uncertainty quantification are:

(i) **Bayesian inference**: Bayesian inference is a statistical inference technique that leverages Bayes' theorem [21] to combine prior knowledge of a model with observed data for analysis. It interprets probabilities as degrees of belief and allows for the estimation and management of uncertainty in the estimates.
(ii) **Monte Carlo simulation:** Monte Carlo simulations predict system outcomes, aiding risk assessment and decision-making [22]. These simulations employ random sampling algorithms to address deterministic problems, distinguishing them from other approaches.
(iii) **Fuzzy systems:** Fuzzy logic is a powerful approach for handling uncertainty in machine learning models. Neuro-fuzzy inference system (ANFIS) is an advanced method integrating fuzzy logic and neural networks to model uncertainty [23]. It combines fuzzy "IF-ELSE" rules and optimal parameters from the fuzzy algorithm to learn non-linear functions. ANFIS's architecture includes five layers for fuzzification, rule generation, normalization, and output generation.
(iv) **Dempster-Shafer's theory (DST):** DST is an extension of Bayesian theory [24]. DST aimed at addressing its limitations, such as the inability to represent ignorance and consider multiple hypotheses. DST, as a theory of evidence, integrates all potential outcomes rather than analyzing individual pieces of evidence.
(v) **Rough set theory (RST):** RST manages uncertainty and inconsistency using an approximation space defined by upper and lower approximations [25]. These approximations can be crisp or fuzzy sets, making RST a fundamental theory in addressing uncertainty.
(vi) **Imprecise probability:** Imprecise probability, a broader concept than traditional probability, allows for estimating uncertainty [26]. Multiple theories exist, including subjective probability and consistent lower prediction, which offer different approaches to modeling imprecise probability.

3.1.1. Related works based on Bayesian inference

In recent years, Bayesian inference has emerged as a versatile and powerful tool for addressing various scientific challenges. This statistical framework allows for the integration of prior knowledge and observed data to make informed estimations and predictions. Furthermore, the Bayesian inference is based on the interpretation of probabilities as degrees of belief. Bayesian rule is used to combine existing information on the a priori known model and unseen data from the sample to be analyzed. This method allows to estimate and effectively manage the inherent uncertainty associated with the estimation process.

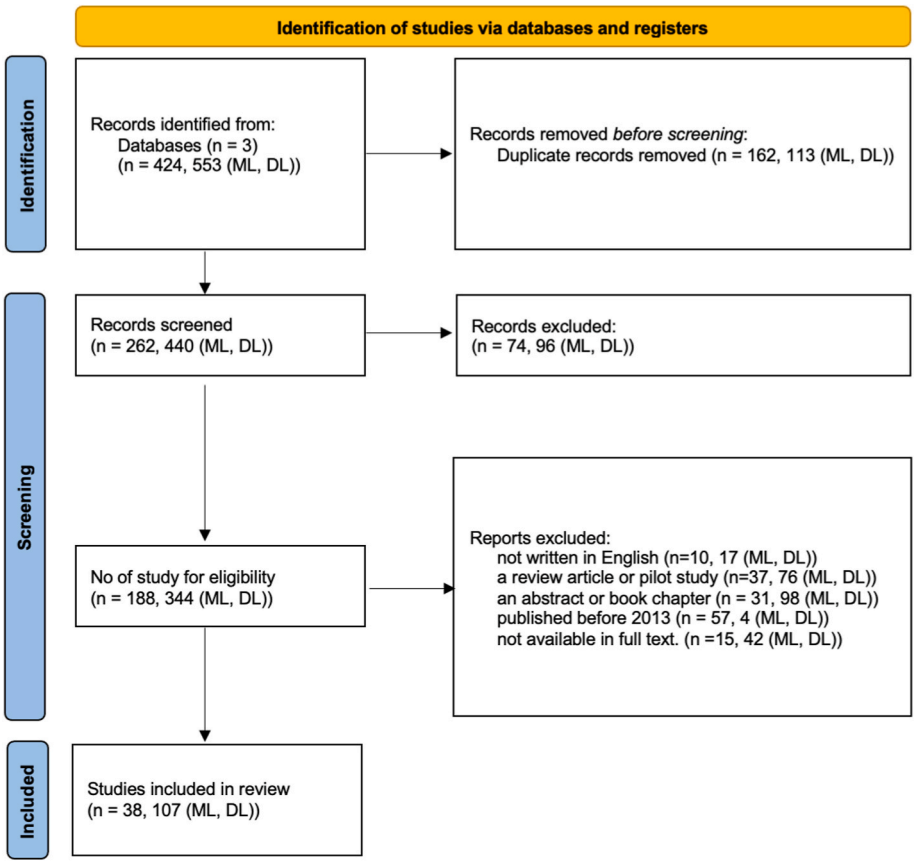

Fig. 3. Selection of relevant articles based on PRISMA guidelines.

Table 2
Summary of the number of papers that employed uncertainty quantification techniques in machine learning frameworks.

Method	N	% of Articles	Reference(s)
Bayesian inference	7	18%	[27–33]
Monte Carlo simulation	6	16%	[33–39]
Fuzzy systems	6	16%	[40–46]
Dempster-Shafer theory	7	18%	[47–53]
Dempster-Shafer theory + Fuzzy logic	5	13%	[54–58]
Rough set theory	4	11%	[59–61,63]
Imprecise probability	3	8%	[64–66]

**N: Number of articles.

Lin et al. [27] developed a framework based on Bayesian inference to estimate the risk factor of nonylphenosis (NPs) exposure in certain foods and environments. The proposed model facilitated the construction of a probabilistic risk estimation framework that considered a population of different age groups and both genders. Zouh et al. [28] developed a model to identify possible artifacts in a reconstructed image, based on the quantification of uncertainty through a Bayesian framework. This application based on Bayesian inference can be used to reconstruct medical images and to estimate the uncertainty associated with the reconstruction itself. Akkoyun et al. [29] demonstrated the effectiveness of Bayesian in estimating the maximum diameter of an abdominal aneurysm from CT images. This estimation enabled the assessment of the aneurysm's growth rate and facilitated the identification of appropriate treatment options. Magnusson et al. [30] proposed an estimation of the principal stratum to assess the effectiveness of the treatment on disability progression in patients with secondary progressive multiple sclerosis (SPSM). A Bayesian inference through Markov chain Monte Carlo (MCMC) methods using the No-U-Turn sampler (NUTS) was used to estimate the principal state. Lipkova et al. [31] presented a Bayesian machine-learning framework to calibrate the mathematical model of glioblastoma tumor growth from multimodal scans. Through a correct inference of tumor density, radiofrequency therapy can be better defined. The Bayesian framework effectively quantified uncertainties in imaging and modeling, allowing for the prediction of patient-specific tumor cell density with credible ranges. Flügge et al. [32] introduced a Bayesian network for the diagnosis of three different kinds of headaches. The study explored three types of inference methods to develop different Bayesian networks for the diagnosis of a brain tumor and three different forms of headache (migraine with/without aura, tension headache and cluster headache). Wang et al. [33] developed a model for assessing the risk factors associated with lung cancer, enabling the development of a medical expenditure model that accounts for data uncertainty through a Bayesian network. By accurately gauging the severity of cancer, the model predicted individual patients' medical expenses, aiding in effective health insurance management.

3.1.2. *Related works based on Monte Carlo simulation*

Another widely used method for dealing with uncertainty is Monte Carlo simulation. Monte Carlo simulations are a class of computational techniques that facilitate the prediction of all conceivable outcomes of a given system, thereby enabling the user to gauge the associated risks and uncertainties prior to making a decision. A distinctive feature of this approach is the use of algorithms that employ a random sampling procedure to tackle deterministic problems.

An example of applying the Monte Carlo simulation tool for uncertainty estimation is presented by Salgado et al. [34], who demonstrated a computer simulation model used to estimate the impact of sugar-sweetened beverages on diabetes and cardiovascular disease. Tsai et al. [35] presented a GPU-based microscopic Monte Carlo simulation tool for the DNA damage caused by ionizing radiations. Specifically,

they presented a GPU-based microscopic Monte Carlo simulation tool for analyzing the DNA damage induced by ionizing radiations. Their work did not revolve around the development of a new chemical or physical model but rather focused on the implementation of a GPU-based model aimed at improving computational cost. Lee et al. [36] used a Monte Carlo simulation to prove the feasibility of the dual-head Compton camera with Si/CZT material as a medical imaging system for the detection of breast cancer. Shih et al. [37] employed the Monte Carlo method to calculate the dose distribution of the blood irradiator, assessing the viability of using MAGAT gel for dose measurements. Unlike traditional dosimeters that necessitate multi-point or plane measurements, the combination of Monte Carlo simulation and polymer gel allowed for the simultaneous acquisition of a 3D dose distribution. Gasparini et al. [38] proposed the Monte Carlo simulation for the evaluation of different analytic models proposed for informative visiting processes in healthcare longitudinal data. This study highlighted the potential for biased regression coefficient estimates within a longitudinal model when an informative visiting process was neglected. Furthermore, various methods proposed in the literature to address this issue were compared and evaluated, with an assessment of the differences in their performance. Lee et al. [39] demonstrated the feasibility of the Monte Carlo simulation to handle the uncertainty of the proton path during the proton therapy. To model the proton beam range monitoring process, they modeled a 3-D PG slit-camera system based on pixelated cadmium zinc telluride (CZT) semiconductor detectors, using TOPAS Monte Carlo simulation.

3.1.3. Related works based on fuzzy systems

Fuzzy logic is a powerful approach for handling uncertainty in machine learning models by accommodating imprecise and ambiguous information. It allows nuanced reasoning and decision-making by assigning membership degrees to different categories. The Adaptive Neuro-Fuzzy Inference System (ANFIS) combines fuzzy logic and neural networks, integrating fuzzy logic with neural networks to model uncertainty [23]. ANFIS adapts its fuzzy inference system's structure and parameters based on input-output training data, enabling accurate inference in complex systems. ANFIS's layered architecture includes input, fuzzification, rule, normalization, and defuzzification stages. Widely used for system modeling, prediction, and control, ANFIS offers a flexible and effective solution for addressing uncertainty.

Castellazi et al. [40] presented several machine learning models combined with unimodal and multimodal MRI features to classify Alzheimer's disease (AD) and vascular dementia (VD). ANFIS proved to be the most effective classifier in distinguishing between AD and VD subjects, achieving the highest performance when using combined tensor imaging (DTI) and genetic testing (GT) features. ANFIS successfully predicted the prevalent underlying disease in 11 out of 15 MXD subjects, resulting in a correct prediction rate of 77.33%. Das et al. [41] proposed a hybrid model called Linguistic Neuro-Fuzzy with Feature Extraction (LNF-FE) to analyze medical data and predict eight different diseases, such as diabetes or breast cancer. The LNF-FE model was developed by incorporating multiple components: expanding input features through fuzzification, assigning linguistic values to these features, performing feature selection using PCA, and using an artificial neural network for prediction. The LNF-FE model exhibited superior performance and achieved better results in comparison to alternative approaches. Vidhya et al. [42] introduced the Modified adaptive neuro-fuzzy inference system (M-ANFIS) for the assessment of various disorders of healthcare. After a data pre-processing phase, a feature selection was performed, and the count of the closed frequent item set (CFI) was estimated. M-ANFIS showed better performance than the other traditional methods, such as SVM. Kaur et al. [43] devised a predictive model for various knee diseases, namely osteoarthritis (OA), rheumatoid arthritis (RA), and osteonecrosis (ON), using a combination of Neuro-Fuzzy and Artificial Neural Network (ANN) techniques. The study involved a comparative analysis of the performance between a fuzzy system and the Adaptive Neuro-Fuzzy Inference System (ANFIS). The results demonstrated the efficacy of the ANFIS approach in accurately predicting knee diseases, providing valuable insights for improved diagnosis and treatment strategies in clinical settings. Liu et al. [44] exploited fuzzy interference logic to develop a decision-making model for prostate cancer detection, analysis, and fusion of medical data and treatment recommendations with risk analysis. De Medeiros et al. [45] developed a fuzzy inference system for supporting medical decisions. The implementation of the Fuzzy Intelligent System demonstrated the potential to create innovative channels for the distribution of medical costs, allowing for accurate assessment of health risks for new patients. Furthermore, its contribution to the medical domain was complemented by increased sales and enhanced hospital marketing efforts, adding value to the overall system. Nguyen et al. [46] developed an integrated system for medical data classification. To be specific, the model consisted of a wavelet transform, for the features extraction, and a type-2 fuzzy logic system for the classification of breast cancers and heart diseases.

3.1.4. Related works based on Dempster-Shafer theory

The Dempster-Shafer theory is a generalization of Bayesian theory [24]. Dempster-Shafer's theory attempts to overcome the limitations of the Bayesian theory, which is unable to describe ignorance and only considers single rows. The Dempster-Shafer theory also known as evidence theory or belief functions theory, provides a powerful mathematical framework for managing uncertainty and combining evidence from multiple sources in machine learning. Introducing belief functions that assign masses of belief to subsets of possibilities, enables a more expressive representation of uncertainty and the ability to handle conflicting evidence. Incorporating the Dempster-Shafer theory into machine learning models enhances their performance and enables more robust and reliable decision-making in uncertain environments.

Buono et al. [47] developed a model based on the Dempster-Shafer theory to generate a diagnosis system for certain skin diseases. After collecting a series of symptoms based on medical knowledge, the authors proposed a set of rules to enable the diagnosis of skin diseases. The Dempster-Shafer method demonstrates its effectiveness in delivering reliable outcomes for skin disease consultation. The results generated by the expert system align with the predetermined rules, thereby confirming the advantage of this method in accurate disease diagnosis. Prameswari et al. [48] introduced the DST to diagnose digestive diseases. Web-based E-diagnostic based on DST provided diagnosis information that was based on symptoms and enables better management of the disease. The results of this study demonstrated that by applying the Dempster-Shafer method for diagnosing digestive disorders in humans, a higher confidence value (70%) was obtained compared to the value obtained (60%) with the Certainty Factor method. Razi et al. [49] addressed the challenge of multi-class motor imagery tasks using a model based on DST. Unlike the traditional common spatial patterns (CSP) method that enables binary classifications, this study focused on analyzing five classes of tasks. To tackle the multi-class problem, a DST-based model was employed, which fused the results of binary classification. Additionally, DST was introduced as a method to handle uncertainty arising from a lack of knowledge in this study. Another interesting application of DST has been proposed by Shi et al. [50], in the context of drug interactions, which can be a key factor in therapeutic decision-making. While descriptions of possible drug interactions exist for many medications, there was no description of the specific interaction analyzed in this study. Building upon this knowledge, the authors presented a model based on local classification (LCM) for predicting drug interactions for new medications. Kang et al. [51] proposed the use of DST in the analysis of the incidence of Clostridium difficile infection (CDI) in hospitals. The proposed model was based on the Gaussian mixture model (GMM) for the generation of the explicit probability criteria to assess the risk factors and the DST for predicting the incidence of infection based on the probability criteria provided by the GMM. A model based on the combination of ambiguity measurement with DST

theory was proposed by Wang et al. [52], for uncertainty management in medical diagnostic decision-making. The ambiguity measure assessed the level of uncertainty for each parameter, enabling the creation of basic probability assignments (BPA) for each parameter. Furthermore, the DST of evidence was employed to aggregate independent evidence into collective evidence, facilitating the ranking of candidate alternatives and identifying the best alternative. Ghesu et al. [53] introduced a model for evaluating medical images, combining uncertainty measurement with probabilistic classification to quantify the system's confidence in its outputs. By employing uncertainty estimation through Dempster-Shafer theory, the model achieved a substantial improvement in accuracy and robustness across different kinds of images, including chest radiographs, abdominal ultrasound image view-classification, and brain metastases detection in brain MR scans.

3.1.5. Related works based on Dempster-Shafer theory and fuzzy logic

Some authors have presented models for uncertainty management based on both fuzzy logic and Dempster-Shafer's theory. For instance, Biswas et al. [54] proposed a model for the enhancement of chest X-ray images based on soft fuzzy sets and the DST approach. The proposed model involved two soft fuzzy sets of the image grey levels. The uncertainty levels of peak intensity and spatial information were handled by fuzzy intervals based on the DST approach. Porebski et al. [55] developed a set of rules for the diagnosis of liver fibrosis based on DST extended for fuzzy focal elements. Utilizing the DST to address knowledge uncertainty caused by incomplete and unbalanced data, the proposed model was successfully developed to support the diagnosis of liver fibrosis. Xiao et al. [56] developed a model to deal with the uncertainty that arises in decision-making. The model integrated belief entropy, fuzzy preference relations, and DST theory to measure and modulate parameter uncertainties while merging independent parameters. The model was validated in a clinical setting, considering four potential diseases: acute dental abscess, migraine, acute sinusitis, and peritonsillar abscess. The proposed method enabled the measurement of parameter uncertainty, and assessment of parameter reliability, and provided insights for clinicians regarding the impact of parameters on decision-making. Ghasemi et al. [57] presented a model for brain segmentation, that was based on the combination of fuzzy inference system and Dempster-Shafer theory (FDSIS). The DST was proposed to handle and reduce uncertainty in MRI segmentation. In the FDSIS algorithm, features were extracted from MRI images, including pixel intensity and spatial information. The fuzzy inference was utilized to construct rules, while the DST was employed for the aggregation phase of the fuzzy inference system. The FDSIS proposed model demonstrated an enhanced accuracy in segmenting both real and simulated MRI images when compared to traditional methods, which generally lack the incorporation of uncertainty estimation and management. Li et al. [58] combined the fuzzy soft set and the Dempster-Shafer theory of evidence for decision-making applied to solving medical diagnosis problems. They used grey relational analysis to calculate the degree of uncertainty of the various parameters, based on which the probability assignment function is obtained. Then, through the Dempster-Shafer rules, all alternatives were aggregated into a collective alternative, whereby they were ranked to obtain the best alternative. The authors demonstrated the superior performance of the model based on fuzzy soft set and Dempster-Shafer theory, surpassing even traditional methods like Feng's method and Naive Bayes' classifier.

3.1.6. Related works based on rough set theory

Rough set theory is a mathematical framework utilized to address uncertainty and inconsistency in data analysis and decision-making. It provides a set of tools for handling imperfect or incomplete information. In the context of uncertainty estimation in machine learning, rough set theory enables the exploration and representation of uncertainty through the definition of upper and lower approximations. It facilitates the identification of uncertain instances and supports attribute reduction, thereby contributing to effective uncertainty management in the learning process.

Acharya et al. [59] developed a combination of cuckoo search and rough set (CRCS) models for knowledge inference from the cardiac disease information system. The objective was to identify which hidden features and knowledge derived from electronic information systems allowed the diagnosis of early cardiac disorders. Clinical data of 603 patients were analyzed, and an initial feature selection using the cuckoo search model yielded eight selected features. These features were then analyzed using rough set data analysis to generate classification rules. The CRCS model exhibits the highest accuracy rate (93%) compared to the rough set model (92%) and the decision tree model (90%), demonstrating its effectiveness in knowledge inference for cardiac diagnosis. Santra et al. [60], showed the use of a lattice of raw knowledge as an information system for the rough set for the design of knowledge for medical expert systems. They applied the proposed model for a simple case study from the domain of low back pain. An innovative metric was proposed to assess the consistency and reliability of rules. The authors demonstrated that the utilization of a lattice of raw knowledge facilitated effective information management in medical systems, surpassing the capabilities of conventional tabular information systems. Bania et al. [61] developed an R-ensemble method for attribute selection by exploiting the rough set theory, demonstrating its superiority over methods already found in the literature. They used a medical dataset, collected from UCI Machine repositories [62], which contained clinical data on Wisconsin breast cancer, lung cancer, diabetes, Indian liver patients, dermatology chronic kidney, and hepatitis. Except for the diabetes dataset, all other datasets exhibited missing values, which were addressed using a k-nearest neighbor (kNN) imputation method. This study aimed to tackle one of the major challenges in the analysis of medical and healthcare data, specifically dealing with datasets that contain missing and redundant information, leading to uncertainty. Jiang et al. [63] showed a novel computational model based on fuzzy mathematics and rough set theory for the assisted diagnosis of sub-health referring to traditional Chinese medicine (TCM). They analyzed original medical records from the First Affiliated Hospital of the Guangzhou University of Chinese Medicine. Comparative analysis with linear models, Naive Bayesian classification, and fuzzy comprehensive evaluation revealed that the novel model achieved higher overall accuracies.

3.1.7. Related works based on imprecise probability

Imprecise probability is a generalization of traditional probability that can be used as a framework in machine learning for handling uncertainty. Unlike traditional probability theory, which assigns precise probabilities to events, imprecise probability allows for the representation of uncertain or ambiguous information by using intervals or sets of probabilities. This approach recognizes that in real-world scenarios, it is often challenging to assign precise probabilities due to limited knowledge or conflicting evidence. In machine learning, imprecise probability provides a flexible and robust framework for uncertainty estimation. It allows for the modeling of uncertain events or variables by considering a range of possible probabilities rather than a single value. This is particularly useful when dealing with incomplete or noisy data, where precise probabilities may be difficult to obtain. For instance, Giustinelli et al. [64] provided empirical evidence on the perception of dementia risk among elderly Americans without dementia and through models on the imprecision of subjective probabilities. Mckenna et al. [65], reported several mathematical models, that highlight the modeling to improve breast cancer treatments, especially chemotherapy, and radiation therapy. The authors demonstrated how mathematical models can provide valuable contributions within the context of breast cancer therapy. In their study, Mahmoud et al. [66], investigated various machine learning models to address uncertainty measures and imprecise probabilities in the diagnosis of medical noisy data. The models proposed in their research were categorized into three groups:

single tree classifiers, ensemble models, and credal decision trees (CDTs). Notably, the credal decision trees outperformed the single tree classifiers, exhibiting higher accuracy, particularly in noisy domains and databases with mostly numerical attributes.

3.2. Uncertainty quantification in deep learning

Table 3 summarizes the distribution of works based on the employed method for uncertainty management in deep learning approaches. Data uncertainty and model uncertainty are both important concepts in data analysis and modeling. Given that various sources of uncertainty may arise in a model, developing effective methods for estimating uncertainty in their prediction is currently a subject of significant interest in the research community [4]. While the data uncertainty is often reflected in the Softmax output of a classification model [67], researchers have extensively investigated *four* main approaches to disentangle and accurately represent model uncertainty from data uncertainty [67,68]. The choice of approach depends on the number and characteristics of the deep neural network being employed [4]:

(i) **Single deterministic methods**: These methods employ a deterministic neural network for uncertainty quantification, relying on a single forward pass to generate predictions without explicitly modeling uncertainty [69].
(ii) **Bayesian methods**: These methods calculate a posterior distribution that captures the uncertainty in the parameter values of the model [70–72]. This distribution is subsequently utilized to quantify the uncertainty in predictions or estimates.
(iii) **Ensemble methods**: These methods leverage the fusion of multiple deterministic networks to enhance model performance and generalization [73]. By combining the predictions of different networks, ensemble approaches enable the generation of more reliable and accurate results surpassing those achieved by individual models alone.
(iv) **Test-time augmentation methods**: These methods involve generating multiple predictions from various augmentations of the primary input data during inference and quantifying uncertainty based on these predictions [74].

When discussing deep learning frameworks, an important aspect is the calibration of the predictor. A predictor is considered well-calibrated when its predictive confidence accurately estimates the actual probability of accuracy [75]. Thus, it is important to ensure that the network is well-calibrated before employing uncertainty methods [4]. There are *three* relevant calibration methods used in healthcare, which depend on the phase they are applied: regularisation [4,76–78] post-processing [79–81], and neural network estimation methods [82,83]. These methods adjust the output probabilities of the model to better match the true probabilities of the data, resulting in more accurate and reliable predictions.

3.2.1. Related works based on single deterministic methods

Signal deterministic methods for uncertainty quantification in deep learning involve deterministic approaches that analyze the characteristics of the model's signals to estimate uncertainty [4]. These methods do not directly model uncertainty as probabilistic distributions but instead focus on analyzing the properties of the model's outputs. Uncertainty can be computed through external methods or internal methods. External methods leverage techniques such as using gradient matrices [84,85], employing additional networks for uncertainty estimation [86], or measuring training data density in the representation space for input data [87]. Internal methods include training prior networks [68], evidential neural networks [88] and using gradient penalties [79]. While they do not provide probabilistic uncertainty estimates, they offer insights into the model's reliability and confidence. These methods can be computationally efficient compared to full probabilistic approaches, making them practical for certain applications.

Ktena et al., [89] developed and trained a convolutional neural network on functional MRI images of the brain. Their objective was to assess the similarity between functional brain networks by measuring the similarity of irregular graphs. Through their proposed method, they achieved a significant improvement of 11.9% in overall classification accuracy. McKinley et al., [90] developed and trained a CNN using MRI images of patients with multiple sclerosis. The authors used best-practice standards to annotate lesions and predict the probability that the network assigns a different label instead of the ground truth. Their approach yielded accuracies of 75% and 85% in accurately distinguishing stable and progressive time points, showcasing the effectiveness of their method. Devries et al., [91] a convolutional neural network and explored six distinct uncertainty estimation techniques to assess uncertainty in the segmentation of skin lesion images. The authors observed that the heteroscedastic classifier neural network yielded the least improvement in results compared to the other uncertainty estimation techniques, which demonstrated comparable performance. Luo et al., [92] introduced a novel deep commensal model for estimating intrinsic uncertainties in cardiac magnetic resonance images. They computed the commensal correlation between direct area estimation and bi-ventricle segmentation, achieving accurate uncertainty estimation through one-time inference based on cross-task output variability. The authors highlight that their proposed method outperforms other approaches in terms of quantification accuracy and optimization results. Ghesu et al., [93] applied a bootstrapping uncertainty measure to their DenseNet model. By employing this recommended uncertainty measurement, the authors found that unwanted training of chest X-ray images could be eliminated, leading to increased robustness and accuracy of the model. Additionally, the method was effective in identifying reader errors. Graham et al. [94] developed and trained a 3-dimensional U-Net model using MRI images of the brain to precisely labeling different regions and sub-regions of the brain. To achieve this, the authors measured cross-entropy uncertainty at progressively smaller sub-regions of the brain. The results showed a dice score of approximately 0.85 for all regions in the uncertainty-aware model, indicating high accuracy in the segmentation task.

Liao et al. [95] developed a DenseNet model to tackle the issue of inter-observer variability in assessing the quality of cardiovascular images obtained through echocardiography. They measured the aleatoric uncertainty by incorporating the variability observed among different experts. The proposed method treated this variability as aleatoric uncertainty and represented it through Laplace or Gaussian distributions in the regression space. The authors observed that their approach resulted in reduced absolute error compared to conventional regression models, as indicated by their findings. Li et al. [96] applied the DistDeepSHAP uncertainty measure to assess the importance of features in autism brain images by employing a SHAP-based deep model. The results indicate that this approach has the potential to identify biomarkers associated with the disease in neuroimaging data. Ye et al., [97] utilized the neurite orientation dispersion and density imaging model and explored the Lasso bootstrap approach for uncertainty estimation of tissue microstructure in brain diffusion magnetic resonance images. The authors observed a meaningful relationship between the proposed uncertainty measures and estimation errors, resulting in the generation of

Table 3
Summary of the number of papers that employed uncertainty quantification techniques in deep learning frameworks.

Method	N	% of Articles	Reference(s)
Single deterministic methods	11	10%	[89–99]
Bayesian methods	68	64%	[105–173]
Ensemble models	14	13%	[176–189]
Test-time augmentation methods	14	13%	[191–204]

**N: Number of articles.

reasonable confidence intervals. Tardy et al. [98] utilized a deep neural network classifier for the classification of mammogram images. The authors estimated network uncertainty using two measurements: subjective logic with softmax predictions and Mahalanobis distance between new and training samples in the embedding space, for three different tasks. They reported that the proposed method allows for the rejection of obvious outliers and improves the area under the curve results by up to 10%. Jensen et al. [99] employed a convolutional neural network for skin image classification. They investigated the use of inter-rater variability sampling during training to improve model calibration. The study demonstrated that the proposed method enhances model calibration, enabling better capture of uncertainty in both samples and labels.

3.2.2. Related works based on Bayesian methods

Bayesian methods involve using different types of stochastic deep neural networks wherein two forward passes of the same data sample generate varying results [4]. In the Bayesian models, the parameters are treated as random variables. During a forward pass, the parameters are sampled from the distribution of data, resulting in stochastic prediction outcomes, where each prediction is based on varying model weights. Bayesian neural networks assume a prior distribution $p(\theta)$ and demonstrate the posterior distribution over the parameter space given by $p(\theta|x, y)$ for the training input pair (x,y). After the estimation of posterior weights, the prediction of an output y^* for the input data x^* may be obtained by performing the Bayesian Mode Averaging or Full Bayesian Analysis [10]. Some types of Bayesian methods include Monte Carlo dropout [101], variational inference [102], sampling [103] and Laplace approximation [104]. The choice of method depends on the specific application and the nature of the data being analyzed.

Leibig et al., [105] employed Bayesian uncertainty measures in combination with various data and deep models to classify fundus images for diabetic retinopathy. The conducted experiments revealed a robust model generalization. Notably, Monte Carlo drop-out outperformed other direct methods, demonstrating its ability to accurately determine and quantify uncertainty. Ozdemir et al., [106] introduced a novel approach where uncertainty measures, specifically predictive mean and standard deviation, were fused with the original image using the Bayesian U-net model. This fusion resulted in the creation of a composite image, which was subsequentially fed into the Bayesian neural network. The authors concluded that incorporating uncertainty measures into the workflow significantly improved prediction accuracy and model confidence. Jungo et al. [107] devised four residual convolutional neural network models with Monte Carlo dropout at full resolution, alongside one model incorporating the conventional weight scaling dropout technique. The position and rate of Monte Carlo dropout were varied for each model, and the performance of these five models was compared to evaluate their effectiveness in uncertainty quantification for brain tumor image segmentation. The authors concluded that informative uncertainty is obtainable by applying the Monte Carlo dropout after each convolutional layer. In a subsequent study [108], the authors developed the U-net model and employed uncertainty techniques such as weighted mean entropy and mean entropy among experts for brain tumor image segmentation. The findings demonstrated that the uncertainty of the model's parameters can be determined by fusing the learned observers' uncertainty with a Monte Carlo-based Bayesian network.

Orlando et al., [109] developed a Bayesian neural network with integrated Monte Carlo dropout to provide epistemic uncertainty feedback. Results showed that the proposed uncertainty estimation inversely corresponds to the model's performance. This highlights its potential use in identifying areas that require corrections in image segmentation. Heo et al., [110] introduced a unique variational attention model that incorporates instance-dependent modeling to capture both data and model uncertainties. The model was validated on six real risk prediction tasks in the healthcare domain, involving physiological signals and images. The authors reported significant improvements achieved by the developed model compared to existing attention models. Adrian et al., [111] integrated the Monte Carlo dropout method with their developed CNN model to estimate uncertainty in multiple sclerosis images. The authors concluded that this technique proves valuable in identifying scans that may require additional examination, as the variance of Monte Carlo dropout samples corresponds to model errors. Roy et al. [112] utilized the Bayesian QuickNAT model and integrated four metrics to assess segmentation uncertainty. The authors highlight that the proposed uncertainty metrics hold promising potential for evaluating the accuracy of segmentation methods in deep models. Herzog et al. [113] combined Bayesian uncertainty techniques and advanced aggregation methods with their Bayesian neural network to achieve highly accurate stroke classification. The authors observed that the integration of Bayesian-based uncertainty methods not only enhanced stroke prediction but also improved the estimation of uncertainty in incorrect patient classification and the detection of uncertain aggregations.

Baumgartner et al. [114] introduced the variational autoencoder model and applied the probabilistic hierarchical segmentation technique on thoracic and prostate images. The results demonstrated that the proposed technique yielded more naturalistic and diverse segmentation of images compared to other related approaches. In a separate study, Raczkowski et al. [115] employed the variational-based dropout measure for uncertainty estimation using the Bayesian neural network in the segmentation of colorectal cancer images. The authors found that the proposed uncertainty measure enhanced the speed of the deep model by approximately 45%, thereby offering a significant computational advantage. Eaton-Rosen et al. [116] conducted a study examining the application of Monte Carlo dropout and M-heads uncertainty measures in the U-net model for calculating predictive intervals during counting tasks in medical imaging. The results indicate that these uncertainty measures are effective in accurately counting histopathological cells and identifying white matter hyperintensity images. Di Scandalea et al. [117] developed a U-net model trained with dice loss and weighted binary cross entropy for segmenting myelin sheath in mice images. They utilized Monte Carlo dropout to estimate uncertainty. The authors highlight that by examining the generated heatmaps from uncertainty estimates, users can identify potential model failures and control uncertainty for more accurate predictions in biomedical applications. Jena et al. [118] employed a Bayesian neural network with a Monte-Carlo uncertainty measure for segmenting brain, cell, and chest radiograph images. The authors concluded that their proposed method improves segmentation quality and calibration, providing more accurate uncertainty estimates compared to existing techniques.

Soberanis-Mukul et al. [119] used a graphical convolutional neural network with Monte Carlo dropout and dice scores as uncertainty measures for segmenting pancreas and spleen images. The authors found that their approach enhances dice scores for both images compared to the original model predictions. Hu et al. [120] utilized the probabilistic U-net model to investigate uncertainty estimation in lung nodule and prostate MRI images. They specifically explored the application of variational dropout. The authors concluded that their approach led to improved predictive uncertainty estimates, enhanced sample accuracy, and increased diversity. Combalia et al., [121] employed a convolutional neural network for the classification of skin lesion images and applied the Monte Carlo dropout uncertainty estimation method. To quantify predictive uncertainty, the authors employed metrics such as entropy, variance, and Bhattacharyya coefficient between distributions. The results indicate the successful utilization of uncertainty metrics in detecting challenging and out-of-distribution samples. Toledo-Cortes et al. [122] developed a hybrid deep learning Gaussian process model for the classification of diabetic retinopathy. In addition to predicting the mean value, the authors also computed the standard deviation as a measure of prediction uncertainty. They found that the proposed model outperformed the original deep learning model and enabled uncertainty analysis. Laves et al. [123] estimated predictive uncertainty using

variational Bayesian inference with Monte-Carlo dropout for regression tasks on medical image datasets. Their findings highlighted that well-calibrated uncertainty in regression tasks enables the elimination of unreliable predictions and the identification of out-of-distribution samples.

In another study by Hu et al. [124], a CNN was trained on PET and CT images for the diagnosis of rare lymphoma. The authors incorporated zone-based uncertainty estimates based on the Monte Carlo dropout technique. The reported sensitivity of the model was approximately 75%, indicating its effectiveness in detecting the target condition. Nair et al. [125] developed a CNN for detecting multiple sclerosis lesions using MRI images from patients with worsening remitting multiple sclerosis. They employed Monte Carlo dropout to approximate probability distributions and subsequently measured variance, predictive entropy, and mutual information. The proposed method achieved a true positive rate of 0.8 and a false detection rate of 0.2, demonstrating its potential for accurate lesion detection. Kwon et al. [126] utilized a Bayesian neural network with predictive uncertainty, which allowed for the decomposition of uncertainty into aleatoric and epistemic components. The authors applied this technique to segment ischemic stroke and retinal images and concluded that it provided a deeper understanding of point predictions. Selvan et al. [127] developed a unique conditional variational autoencoder called conditional Normalizing Flow (cFlow) to improve the approximation of latent posterior distributions. The performance of their model was evaluated on two medical imaging datasets, demonstrating substantial improvements in both qualitative and quantitative measures compared to state-of-the-art methods. In a study by Seebock et al. [128], the Bayesian U-net model with Monte Carlo dropout was employed to estimate model uncertainty in retinal image segmentation. The authors found that the proposed technique achieved high accuracy in segmenting both healthy and diseased retinal images. Hiasa et al. [129] employed the Bayesian U-net CNN model along with Monte Carlo dropout and dice scoring for uncertainty estimation in muscle CT image segmentation. They discovered a relationship between high uncertainty pixels and segmentation failure, enabling patient-specific analysis of muscles. Xia et al. [130] implemented a Bayesian model with uncertainty-aware multi-view training on pancreas and liver tumor images. The authors concluded that applying multi-view co-training on 2D models yielded promising results. Marc et al. [131] investigated the integration of reversible blocks into the PHiSeg architecture for image segmentation. The authors reported that the recommended method required less memory compared to not using reversible blocks while maintaining comparable segmentation accuracy.

Wickstrom et al. [132] used a CNN with a Monte Carlo dropout backpropagation algorithm to determine the uncertainty in input feature importance. The authors demonstrated that their proposed method effectively models uncertainty in input feature importance, showing significant contrasts between correct and incorrect predictions. Carneiro et al. [133] used a DenseNet model to investigate uncertainty estimation and confidence calibration in the classification of colorectal polyps. They explored both Bayesian and non-Bayesian inference methods, using entropy as an uncertainty measure. The study demonstrated that employing Bayesian methods to determine classification entropy or variance resulted in an accuracy of approximately 76%. Li et al. [134] developed a CNN model and compared three Monte-Carlo dropout methods, evaluating metrics such as negative log-likelihood, and expected calibration error. The authors found that the proposed method of region acquisition, as opposed to full region acquisition, led to better calibration of the model regardless of the uncertainty measure used. Quan et al. [135] proposed a deep CNN model and explored Bayesian uncertainty estimates and ensemble semi-supervised learning for correcting noisy labels in upper gastrointestinal images. The proposed method effectively improved recognition accuracy for both authentic and noisy clinical data.

Wang et al. [136] employed a unique approach by implementing a Bayesian teacher-student deep model with Monte Carlo dropout to estimate segmentation and feature uncertainty in atrial MRI and kidney CT scan images. The authors found that their proposed method outperformed existing semi-supervised uncertainty estimates on both datasets, demonstrating its effectiveness in uncertainty estimation. Bian et al. [137] combined a segmentation network with a Conditional Variational Autoencoder (CVAE) for uncertainty estimation, using the variance of the network's output as a measure of uncertainty. They proposed an Uncertainty-aware Cross Entropy (UCE) loss to leverage uncertainty information and improve segmentation performance in highly uncertain regions. The findings demonstrated that the proposed method outperformed existing methods for unsupervised domain adaptation tasks. Tanno et al. [138] combined a noise model with Bayesian inference for uncertainty estimation in brain tumor image datasets. Their results demonstrated that measuring uncertainty improved prediction performance and enabled the detection of predictive failures. Additionally, the decomposition of predictive uncertainty provided high-quality explanations for model performance. Thiagarajan et al. [139] employed and compared Bayesian-based and transfer learning CNN models for uncertainty estimation in breast histopathology images. The findings showed that the Bayesian CNN model outperformed existing models and was useful in explaining uncertainties in histological images. Ghosal et al. [140] introduced two innovative techniques, Monte Carlo DropWeight and Bayesian Residual UNet, specifically designed for estimating aleatory and epistemic uncertainty. By employing these methods, the authors were able to accurately estimate uncertainty, significantly boosting the confidence of clinicians in the field of semantic segmentation. Edupuganti et al. [141] employed variational autoencoders and convolutional neural network models to quantify uncertainty in MRI segmentation of knee images. They utilized Monte Carlo sampling to create a posterior of image pixel variance maps and achieved a SURE-MSE (Stein's Unbiased Risk Estimator) value of 0.97 for 2-fold under-sampling.

Valliuddin et al. [142] utilized a probabilistic U-Net model to perform density modeling on thoracic computed tomography and endoscopic polyp images. They employed a probabilistic segmentation model to learn aleatoric uncertainty as a distribution of possible annotations. The authors concluded that this approach improved predictive performance by up to 14% in modeling uncertainty. Teng et al. [143] employed a deep generative model with recurrent neural networks and trained it using clinical, imaging, genetic, and biochemical markers to investigate the progression of Alzheimer's and Parkinson's disease. The model, incorporating internal stochastic components, achieved good accuracy of 98.1% and 79.7% for Alzheimer's and Parkinson's disease, respectively. Wang et al. [144] applied a multi-instance learning approach for the classification of diabetic macular edema using optical coherence tomography images. They quantified uncertainty by measuring the mean and standard deviation of probabilistic predictions, resulting in an accuracy of approximately 95%. Zhang et al. [145] explored deep neural networks, random forest classifiers, and the light gradient boosting model for toxicity prediction in chemical compounds. They employed conformal prediction with user-defined significance levels to quantify prediction uncertainty, obtaining an average AUC of 0.734. Vranken et al. [146] utilized deep Residual Inception Networks to investigate aleatoric and epistemic uncertainties in 12-lead electrocardiogram signals. The authors concluded that variational inference with Bayesian decomposition and ensemble with auxiliary output performed the best, but high uncertainty in deep neural network-based ECG signal classification correlated with lower diagnostic agreement compared to the interpretation of cardiologists. Sieradzki et al. [147] employed a deep generative model with recurrent neural networks for compound bioactivity prediction. They utilized dropout-based uncertainty estimation by passing test samples through the network with weight dropout, measuring uncertainty from variance. The proposed method achieved precision values between 0.0004 and 0.0007.

Natekar et al. [148] developed convolutional neural network models

for brain MRI image classification to detect brain tumors. They computed the mean of the variance in a predicted posterior distribution obtained by running. Sedghi et al., [149] employed a CNN to assess model agreement in brain image registration. The authors computed the variance in displacements for various brain MRI images. This approach facilitated the estimation of local registration uncertainty, which helps identify areas where the two images may not align well and provide information to end-users about the registration quality. Norouzi et al. [150] employed fully convolutional neural networks for cardiac image segmentation and computed model uncertainty by estimating the variance of the model's output. They further enhanced segmentation accuracy using conditional random fields and assessed the proposed approach with three different metrics. The authors emphasized the incorporation of new techniques and the successful integration of simple ideas with deep neural networks. Filos et al. [151] conducted a systematic study comparing various uncertainty estimation methods using Bayesian deep learning techniques for diabetic retinopathy classification. Their research emphasized the importance of systematic comparisons to demonstrate the efficacy of Bayesian deep learning techniques on large-scale problems. Ghoshal et al. [152] utilized a Monte-Carlo Dropweights Bayesian Convolutional Neural Networks (BCNN) model to estimate uncertainty in predictions of deep learning models applied to chest X-ray images of patients with COVID-19. Their results revealed a correlation between uncertainty and prediction accuracy. Dolezal et al. [153] developed deep convolutional neural network models for the classification of lung adenocarcinoma and squamous cell carcinoma in out-of-distribution digital histopathological data. They estimated slide-level uncertainty for whole slide images by applying uncertainty thresholding to generalize the handling of out-of-distribution data. The findings of the study corroborated that high-confidence predictions outperform those without uncertainty, and uncertainty thresholding is a reliable approach for making high-confidence predictions in lung adenocarcinoma and squamous cell carcinoma out-of-distribution data. Mensah et al. [154] uniquely employed Bayesian capsule networks for uncertainty estimation on computer vision and chest X-ray image datasets using mean-field variational inference. They highlighted the transparency, credibility, reliability, and interpretability of Bayesian capsule networks in gaining the confidence of industry partners. Mazoure et al. [155] developed a distinctive web server for deep uncertainty estimation of skin lesion images, specifically for skin cancer detection. They compared the means and variances from new and traditional convolutional neural network models. The findings established that the proposed method outperforms other supervised, self-supervised, and uncertainty estimation techniques, making it the best-performing approach in skin cancer detection. Jahmunah et al. [156] employed the deep DenseNet model to estimate predictive entropy for the misclassification of normal and myocardial infarction ECG signals. Based on the obtained results, the authors asserted that the proposed model is reliable, trustworthy, and confident in the diagnostic information it provides. Therefore, it holds great potential for utilization in healthcare applications. Stoean et al. [157] investigated the use of Monte Carlo dropout within the DL structure to automatically identify indicators of spinocerebellar ataxia type 2 from saccadic samples obtained from electrooculograms. Unlike the typical integration of this specific dropout method in deep neural networks, the researchers used the uncertainty derived from validation samples to construct a decision tree at the patient register level. This decision tree, constructed from uncertainty estimates, achieved a classification accuracy of 81.18% in distinguishing between control, presymptomatic, and symptomatic classes. Guo et al. [158] proposed a technique to enhance multi-class segmentation of cardiac MRI by combining CNNs with interpretable machine learning algorithms. This approach demonstrated significant improvement over traditional CNN segmentation. Evaluations were performed on two distinct cardiac MRI datasets representing various cardiovascular pathologies, with the proposed model exhibiting increased segmentation accuracy and reduced variability. In a separate study, Da Silvia et al. [159] presented a Monte Carlo method-based approach to analyze the performance of measurement systems during design phases to improve their quality. They focused on a simulated electrocardiogram system, using measurement uncertainty as a performance parameter during the design process. The Monte Carlo method enabled the identification of the primary source of ECG measurement uncertainty, aiming for better characterization of the metrological behavior of ECG measurements. Nasir et al. [160] introduced a model for the early prediction of type 2 diabetes mellitus (T2DM) using real-world electronic health record (EHR) data, which included historical diagnoses, patient vitals, and demographic information. By employing Monte Carlo dropout for uncertainty estimation, the proposed model demonstrated a 1.6% accuracy improvement compared to baseline techniques. Abdar et al. [161] developed a novel deep learning model called UncertaintyFuseNet, specifically designed for the accurate classification of large CT scan and X-ray image datasets in COVID-19 cases. The model integrated the Ensemble Monte Carlo Dropout (EMCD) technique, which effectively estimated uncertainty during the learning process. The experimental results showcased the model's efficacy, with impressive prediction accuracies of 99.08% for CT scan datasets and 96.35% for X-ray datasets. Additionally, UncertaintyFuseNet displayed robustness to noise and reliable performance when applied to unseen data. MacDonald et al. [162] conducted a comparative analysis of three approximate Bayesian deep learning models for predicting cancer of unknown primary origin, using three RNA-seq datasets consisting of 10,968 samples across 57 cancer types. The study demonstrated that Bayesian deep learning is a promising approach for generalizing uncertainty, thereby improving the performance, transparency, and safety of deep learning models in real-world applications. Farooq et al. [163] proposed a residual-attention-based, uncertainty--guided mean teacher framework that incorporated residual and attention blocks for breast cancer detection. The quantitative and qualitative findings showed that the proposed framework outperformed state-of-the-art techniques and surpassed existing methods for breast ultrasound mass segmentation. The study also highlighted the potential of including additional unlabeled data to enhance breast tumor segmentation performance. Abdar et al. [164] proposed a simple, yet novel, hierarchical attentive multilevel feature fusion model that leveraged uncertainty quantification during predictions in the classification task. By integrating dropout and Bayesian inference techniques, they effectively enhanced the performance in terms of accuracy, recall, and precision for classification in OCT, lung CT, and chest X-ray. Zakeri et al. [165] introduced DragNet, an unsupervised statistical motion model with Bayesian uncertainty quantification for generating high temporal resolution image sequences from a single reference frame. DragNet offered analytical spatiotemporal uncertainty estimation at the pixel level in a cardiac cycle. Abdar et al. [166] proposed a Binarized Multi-Gate Mixture of Bayesian Experts (MoBE) ensemble technique for accurate cardiac syndrome X (CSX) classification, using uncertainty strategy with Bayesian neural networks (BNNs) and dropout Monte Carlo for decision uncertainty quantification. Achieved impressive 85% accuracy on Tehran Heart Center's CSX dataset. Tanno et al. [167] proposed Bayesian inference-based methods for capturing uncertainty in medical image enhancement using deep learning. A spatial map of predictive uncertainty over output image enabled subject-specific and voxel-wise reliability assessment, demonstrating benefits in enhancing system safety for diffusion MRI super-resolution through Image Quality Transfer (IQT). Wang et al. [168] introduced a Bayesian inference approach for CT image segmentation of cochlear structures. The framework balanced shape and appearance information using likelihood appearance and prior label probabilities based on a generic shape function, showing promising results on multiple datasets. Corrado et al. [169] utilized Bayesian probabilistic methods to estimate left atrium anatomy from Cardiac Magnetic Resonance images. The proposed model quantified uncertain left atrial shape, accounting for imaging artifacts, and assessed its impact on left atrial activation time simulations. The

authors demonstrated that quantifying the uncertainty of the shape impacts the simulation of cardiac activation in the left atrium. Dhamala et al. [170] used Direct Markov Chain Monte Carlo for uncertainty estimation in personalized modeling with small-sized datasets, enhancing clinical decision-making reliability. The framework was evaluated in cardiac electrophysiological modeling using synthetic and real data experiments, revealing valuable parameter uncertainty insights through efficient surrogate modeling integration. Chen et al. [171] introduced TransMorph, a cutting-edge model for unsupervised deformable image registration. Distinguished from traditional approaches, TransMorph leveraged the Transformer architecture and incorporates Bayesian deep learning to estimate deformation uncertainty without compromising registration performance. The model's validation on brain MRI images and phantom-to-CT images showcased superior accuracy compared to conventional methods. Abdullah et al. [172] proposed a study to assess uncertainty in multi-layer perceptron (MLP) Mixer models and CNN models for small datasets using Bayesian Deep Learning (BDL) techniques. Their results showed that BDL significantly improved MLP-Mixer performance by 9.2%–17.4% across various models. Dolezal et al. [173] introduced *Slideflow*, a versatile deep learning library for histopathologic image processing and visualization. This library integrates uncertainty estimation using Monte Carlo Dropout into a variety of deep learning models for stain normalization, augmentation, and classification.

3.2.3. Related works based on ensemble methods

Ensemble methods involve the combination of many different deterministic networks during model inference. Hence, the prediction from an ensemble model is based on diverse predictions obtained from the different networks. Using combined effects among different networks, researchers have found that a group of networks tends to make better decisions than a single network, leading to improved model generalization [4]. Ensemble models may be trained using weight sharing [174], reducing numbers [6], and other various strategies like data shuffling or boosting [175].

In a recent study by Jungo et al., [176], subject-wise uncertainty measures were compared against five other uncertainty measures, including ensemble models, for brain and skin lesion image segmentation. The authors discovered that while existing uncertainty measures demonstrate good calibration at the data level, they are not well-calibrated at the subject level. Hence, subject-wise uncertainty estimates are crucial measures for accurate segmentation. McClure et al., [177] proposed the MeshNet architecture combined with the distributed weight consolidation technique to train independent structural MRI datasets. The findings revealed that the distributed weight consolidation measure improved the performance of each independent test while maintaining model generalization, surpassing the standard ensemble model. Wu et al., [178] introduced the deep Dirichlet mixture model to generate point estimates and credible intervals from learned distributions for evaluating uncertainties in Alzheimer's disease classification probability. The authors discussed the usefulness of the proposed model in predicting uncertainties for multiclass classification problems.

Linmans et al., [179] conducted a comparative analysis between the performances of the Multi-head convolutional neural network model combined with meta-loss functions, and those of the Monte Carlo dropout and deep ensemble methods, for estimating predictive uncertainty on out-of-distribution lymph node tissue images. The authors concluded, based on the results, that the multi-head convolutional neural network outperformed both Monte Carlo dropout and deep ensembles. Liang et al. [180] developed and trained four different types of CNN models using diverse datasets consisting of head CT, mammography, chest x-ray, and histological images. Instead of using the cross-entropy loss function, they introduced an auxiliary loss term that captures the difference between predicted confidence and accuracy for classification tasks, aiming to quantify model calibration error. The authors discussed that their proposed approach significantly reduces calibration error across various models and datasets. Hoebal et al. [181] compared the performance of traditional U-Net, U-Net with Monte Carlo dropout, and Deep Ensemble in segmenting nodules in CT images. The Deep Ensemble method showed slightly better results compared to the Monte Carlo dropout. The authors concluded that incorporating uncertainty information provides a means to assess segmentation quality automatically, even without access to ground truth. Mehrtash et al., [182] employed a fully convolutional neural network (FCN) along with model ensembling to calibrate model confidence. They conducted a comparison of results using both dice and cross-entropy losses. The authors found that employing model ensembling successfully calibrated the confidence of fully convolutional neural networks trained with the dice loss function. Dahal et al. [183] investigated three uncertainty measures and utilized four metrics on the ResNet model for cardiac ultrasound image segmentation. The results demonstrated that uncertainty estimation effectively identified and rejected low-quality images, leading to enhanced segmentation outcomes. The study employed three ensembling-based uncertainty models quantified using four different metrics. Chiou et al. [184] utilized an encoder-decoder network combined with a CycleGAN-based approach for uncertainty estimation in prostate image segmentation. The findings established that the proposed method improved image representations in prostate image segmentation, particularly for cancer characterization. Cao et al., [185] developed a temporal ensembling segmentation model to segment and classify masses in breast ultrasound. An uncertainty-aware unsupervised loss was also integrated into their model. Thanks to this approach, the authors obtained a pixel-wise accuracy of about 99%. Qin et al., [186] employed a CNN to estimate brain and cerebrospinal fluid intracellular volume. The authors trained an ensemble of deep models and calculated the variance in the combined results. The findings demonstrated significant relationships between estimation uncertainty and error across all measurements. Singh et al. [187] developed the Bayesian Multi-ResUNet model for the segmentation and classification of skin lesion images. The authors thoroughly investigated the effectiveness of two techniques: Monte-Carlo dropout and test time augmentation. Their findings revealed that the recommended approach not only showcases the robustness of the model but also enhances its transparency and confidence. Guo et al. [188] introduced a globally optimal label fusion algorithm and an uncertainty-guided, coupled continuous kernel cut algorithm for deep learning with shape priors. These were integrated into a deep learning ensemble algorithm designed for left ventricle segmentation and functional measurements in short-axis cardiac cine MRI. Remarkably, their model exhibited outstanding performance even when trained on small datasets (5–10 subjects) and with sparse annotations. Buddenkotte et al. [189] introduced an efficient model to calibrate deep learning ensembles for accurate classification probability approximation in medical image segmentation of ovarian or kidney tumors. The approach was successfully validated with complex segmentation tasks using large 3D networks, showing that the generated heatmaps outperformed traditional methods in approximating classification probability.

3.2.4. Related works based on test-time augmentation methods

Test-time data augmentation methods involve predicting and quantifying uncertainty at inference based on multiple predictions generated from various augmentations of the primary input data. Typically, multiple test data are created from each input data by applying data augmentation methods; then, the entire set of test data is used to calculate the predictive distribution for the estimation of uncertainty [4]. Greedy policy search [190] is an example of an augmentation policy where each stage of the search selects a sub-policy that provides the most significant improvement in the ensemble predictions, which is added to the existing policy. These methods can improve model robustness and generalization by generating a diverse set of augmented data for testing and prediction.

In their study, Wang et al., [191] introduced a unique approach by

using a CNN model with a bounding box for the segmentation of fetal and brain tumor images. They further explored scribble-based segmentation and image-specific fine-tuning during testing. The authors concluded that the proposed fine-tuning technique significantly improved segmentation accuracy while reducing user time and interactions required for the process. Ayhan et al., [192] developed a CNN model and incorporated a conventional geometric and color transformation technique as an uncertainty measure during testing on fundus images. The aim was to analyze the variations in the network's output. Based on their findings, the authors reported that their test-time augmentation approach provides valuable approximations for predicting uncertainties in deep models. Wang et al. [193] investigated aleatoric and epistemic uncertainties by incorporating test-time augmentation and test-time dropout methods into their CNN model. The authors' analysis revealed that the aleatoric uncertainty estimation technique yielded superior advantages compared to the test-time dropout technique. Specifically, it effectively mitigated the issue of overconfident predictions, resulting in more accurate and reliable uncertainty estimates. Zhang et al. [194] introduced a novel method for MRI reconstruction, where measurements are dynamically selected, and the prediction is iteratively refined during inference to achieve optimal reconstruction. The authors found that this technique effectively reduces reconstruction uncertainty in the resulting images. Athanasiadis et al., [195] developed a generative adversarial network to explore the relationship between visual and audio emotional expressions. They employed conformal prediction to obtain calibration error and confidence values during testing. The authors reported an approximately 2% increase in classification accuracy on two public datasets. In a separate study, Ayhan et al. [196] developed and trained a convolutional neural network using fundus images for diagnosing diabetic retinopathy. They calculated the variance using entropy as a measure of the distribution of predicted probabilities. The reported accuracy ranged from a commendable 96%–98%, highlighting the effectiveness of their approach in achieving accurate diagnoses. Araujo et al., [197] used convolutional batch normalization blocks and max-pooling layers to assess the severity of diabetic retinopathy in retinal images. The authors employed Cohen's kappa statistics to estimate the model's predictions at different uncertainty threshold levels by calculating the variance in image-wise retinopathy grade probabilities. They concluded that the best results were achieved using the quadratic-weighted Cohen's kappa, ranging from 0.71 to 0.84.

Abdar et al. [198] developed a hybrid deep model for skin cancer image classification. They explored three uncertainty metrics, including Monte Carlo dropout and Deep Ensemble. The results demonstrated that the proposed model achieved the highest accuracy of approximately 91% and showed potential for effective use in various stages of medical image analysis. Scalia et al. [199] employed graph convolutional neural networks for predicting molecular properties. The authors quantified prediction uncertainty using Monte Carlo dropout, deep ensembles, and bootstrapping methods on four datasets. The deep ensemble consistently outperformed the other techniques. Dong et al. [200] introduced a novel deep neural network model called RCoNetks, designed for COVID-19 detection in chest X-ray (CXR) images. The model generates both the final diagnosis and uncertainty estimation, and it has been tested on both the original dataset as well as a corrupted dataset containing varying percentages of fake samples. In the presence of noise within the data, the proposed method displayed superior effectiveness compared to existing approaches. Cortes-Cirianco et al. [201] utilized ensembles of several deep learning models to examine the effectiveness of a substance in inhibiting a biochemical or biological function. They monitored the model parameters during single network optimization and calculated the variability and validation residuals across snapshots to quantify prediction uncertainty. The findings revealed a strong relationship between confidence levels and the percentage of confidence intervals, indicating accurate bioactivity estimation. In a separate study, Cortes-Cirianco et al. [202] utilized deep neural networks and the random forest classifier to explore the effectiveness of a substance in inhibiting a biochemical or biological function. They employed conformal prediction, along with test-time dropout, to compute prediction errors on various prediction combinations. The authors concluded that there existed a robust correlation between confidence levels and error rates in their analysis. KarAzmoudeh et al. [203] introduced Bayesian approximation and ensemble learning techniques as uncertainty quantification methods for classifying breast tumor tissue. They demonstrated that by employing evaluation criteria based on uncertainty estimation, it is possible to determine when to trust the output of a deep neural network. Furthermore, the Bayesian Ensemble model displayed greater reliability in quantifying uncertainty. Graham et al. [204] introduced a model for segmenting colon histology images, utilizing uncertainty quantification during test time by applying random image transformations. They also presented an uncertainty-based score to assess prediction reliability. The model exhibited excellent performance in segmenting both gland lumen and gland object across datasets from different centers.

4. Discussion

This paper discusses the importance of incorporating uncertainty estimation techniques in healthcare applications of machine and deep learning models. While Explainable AI (XAI) is a growing area of research, relying solely on XAI techniques cannot guarantee the reliability of model decisions. To promote safe decision-making in the medical domain, it is crucial to present uncertainty estimates in AI systems. Fig. 4 shows the main advantages of using UQ in AI models.

4.1. Uncertainty in machine learning frameworks

In machine learning, several approaches are being investigated to address decision-making under uncertainty, including Bayesian networks, Fuzzy logic, Monte Carlo simulation, and Dempster-Shafer theory [205]. Bayesian networks rely on the concept of conditional independence to compute values of the joint distribution based on random variables in a specific domain. On the other hand, Dempster--Shafer's theory quantitatively evaluates uncertainties through subjective assessments of statement reliability by experts. Fuzzy logic assigns values to elements using membership functions that represent their degree of belongingness to a fuzzy set, with subjective probability distributions assigned to these fuzzy sets [205]. When it comes to medical decisions, Bayesian networks, and fuzzy logic are preferred due to their ability to represent medical knowledge in a structured manner and efficiently utilize prior probabilities for problem-solving [205]. These concepts are summarized in Table 2, which highlights Bayesian models and Dempster-Shafer theory as the primary methods for uncertainty estimation in healthcare using machine learning techniques. This suggests that these approaches have proven effective in handling uncertainty in medical data and improving prediction accuracy. Additionally, uncertainty techniques utilizing machine learning models have primarily been applied to neurological systems, followed by thoracic systems (as cardiac systems), medical data, and other organs (with breast cancer detection being the most extensively studied) (Fig. 5). Uncertainty plays a significant role in machine learning, particularly in the analysis of clinical data (62%) and biomedical images (24%). Flügge et al. [32] investigated diagnostic inference when faced with uncertainty using Bayesian networks. Their model was tested on real-world medical history data, and the information derived from Bayesian networks can be applied beyond the mere determination of diagnostic probabilities for a given medical history. On the other hand, Lipkova et al. [31] demonstrated a Bayesian machine learning framework that utilizes high-resolution MRI scans and highly specific FET-PET metabolic maps to design personalized radiotherapy plans and estimate tumor cell density in patients with glioblastoma. This approach offers a promising avenue for individualized treatment planning and could lead to

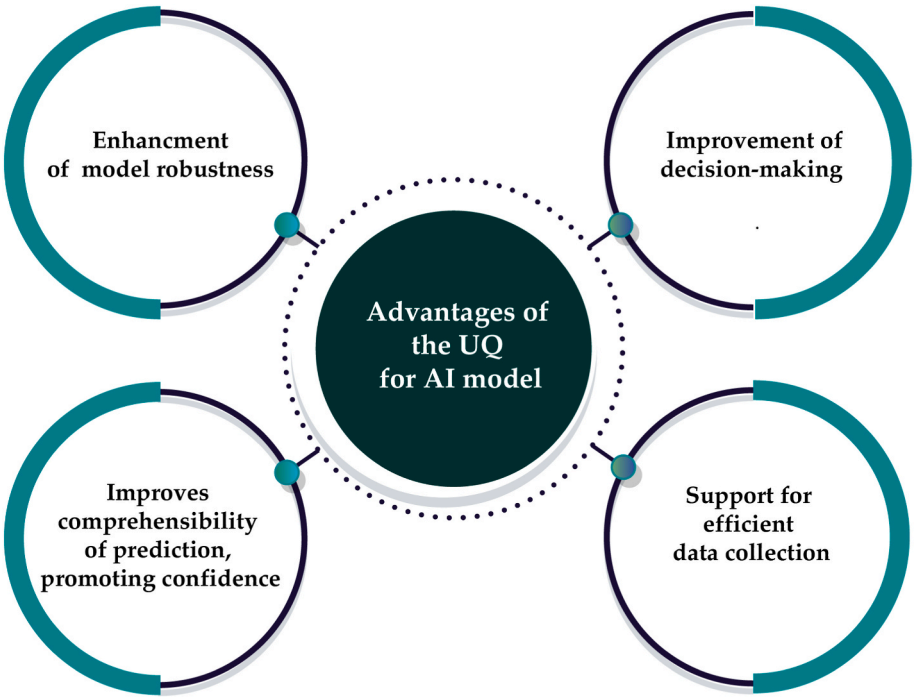

Fig. 4. The main advantages of using UQ in AI models.

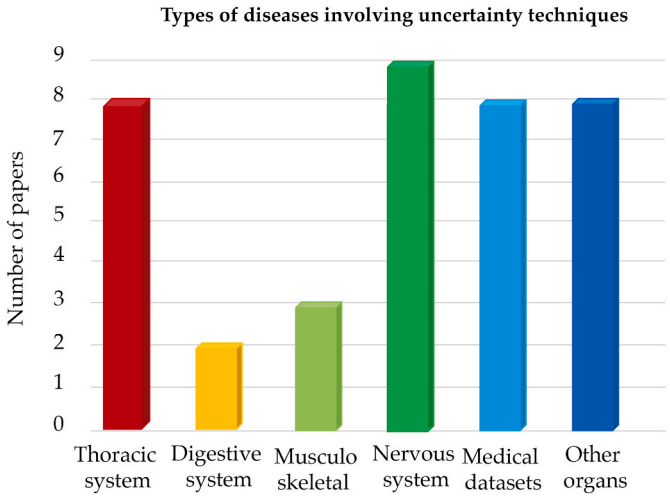

Fig. 5. Types of diseases most prevalently studied involving uncertainty techniques using machine learning models in healthcare (Table 2). The medical dataset represents works that utilize different combined datasets or non-specific datasets, such as EHR.

improved clinical outcomes. Razi et al. [49] is the only study that investigates uncertainty in signal processing models, demonstrating a new method for classifying motor imagery tasks based on Dempster-Shafer's theory.

Integrating uncertainty measurement into machine learning frameworks offers multiple benefits. It improves decision-making by providing insights into prediction confidence. Uncertainty estimation enhances model robustness, detecting out-of-distribution inputs. It facilitates interpretability, building trust and allowing experts to validate model decisions. Additionally, uncertainty-aware frameworks support efficient data acquisition strategies. Overall, it empowers users with reliable predictions, enhances model robustness, promotes interpretability, and supports efficient data collection.

4.2. *Uncertainty in deep learning frameworks*

Fig. 6 shows the different types of images studied for uncertainty techniques used in healthcare based on deep learning frameworks. The analysis reveals that brain, eye, and skin images have been the most extensively studied in the past decade, followed by chest, cardiac, and breast images. However, limited instances of research exist for liver, spleen, gastrointestinal tract, muscle, audio-visual, and cell membrane images, possibly due to challenges related to biological variability, imaging modality, and expert annotation [206]. This variability may explain why brain, eye, and skin images are more commonly associated with uncertainty techniques compared to other medical images. Non-imaging data, such as physiological signals and the bioactivity of proteins, have received limited attention in the literature. Only a few studies, such as Heo et al. [110] examining data and model uncertainties using multiple physiological signals, and Jahmunah et al. [156]

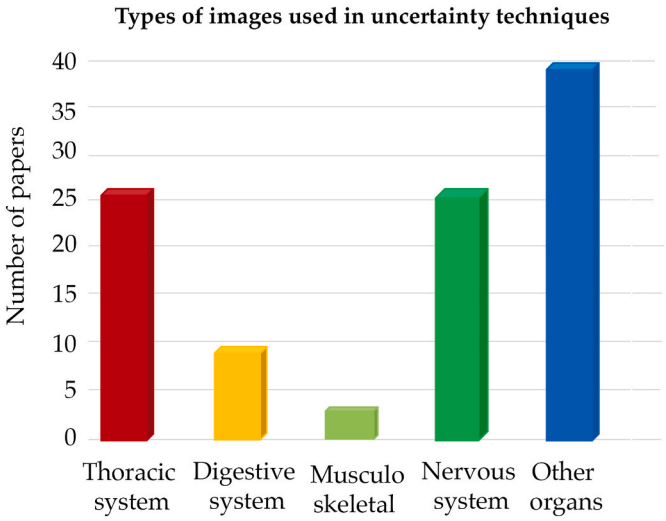

Fig. 6. Bar graph representing the different types of images used in Table 3.

investigating model uncertainty using ECG signals, are mentioned in this review paper. These findings indicate that while uncertainty techniques have been extensively explored in the context of medical images, their application to non-imaging data is still emerging in the healthcare domain. Indeed, only a few studies have focused on studying uncertainty applied to physiological signals, such as in the case of ECG signals [156].

Fig. 7 presents a pie chart illustrating the use of deep learning models with uncertainty techniques in healthcare. The analysis reveals that approximately half of the studies incorporated uncertainty techniques into convolutional neural networks (CNN), followed by Bayesian-based deep learning models. Deep CNN models are effective in learning useful representations of images and structured data [207] while Bayesian neural networks are effective in describing model uncertainties while requiring low memory consumption [4]. In contrast, models such as autoencoders, and ensemble models were used to a lesser extent. Ensemble methods do not effectively describe model uncertainties, require training many networks, and incur high computational effort and memory consumption [4]. Autoencoders employ the axis-aligned Gaussian as the latent distribution, which may be disadvantageous when estimating complex latent posterior distribution [127] in uncertainty techniques. Consequently, CNN and Bayesian deep models are more prominently utilized with uncertainty techniques in image-related applications, due to their inherent strengths and advantages over autoencoders and ensembles, as discussed.

4.3. Key papers and techniques in the field of uncertainty quantification

In the realm of machine learning, all methods for uncertainty quantification are equally represented (as shown in Table 2). However, in the domain of deep learning, Bayesian methods are the most widely used, with 60% of the papers (n = 68) identified in our review employing Bayesian methods (as shown in Table 3). Fig. 8 illustrates the various types of data employed alongside Bayesian methods, the predominant approach for UQ. It emphasizes that thoracic system data is the most prevalent, followed by nervous system data. Among these Bayesian methods, the MC dropout technique is the most popular, for several reasons. Firstly, the implementation of MC dropout is relatively straightforward, and unlike other techniques for quantifying uncertainty, it only requires enabling dropout layers during test-time to obtain uncertainty maps. This makes it a convenient and accessible method for many researchers and practitioners. Secondly, MC dropout is highly flexible and can be implemented in most deep neural networks simply by adding dropout layers within the architecture. This means that it can be used with a wide range of models and architectures, making it a versatile

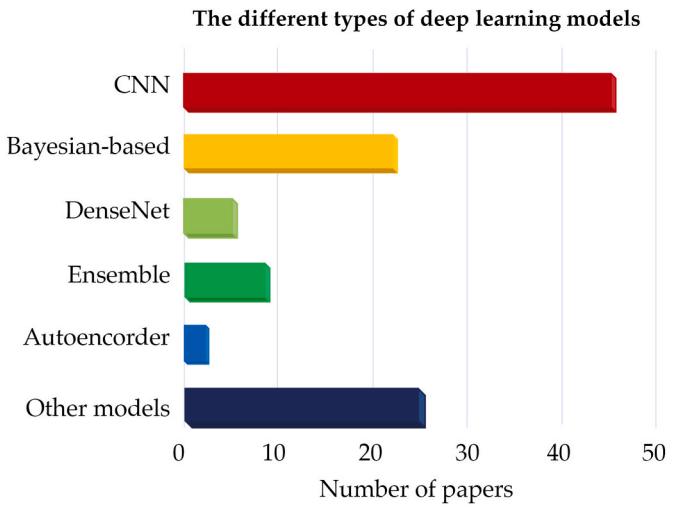

Fig. 7. Pie chart representing the different types of deep learning models used to model uncertainty in deep learning frameworks.

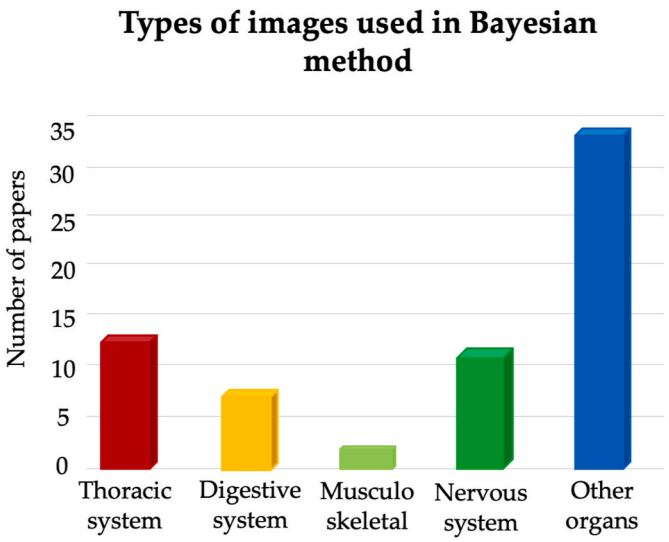

Fig. 8. Bar graph representing the different types of images used with the Bayesian methods.

tool for uncertainty quantification. Finally, once uncertainty maps are obtained using MC dropout, they can be used to customize the pipeline by imposing fixed or adaptive thresholds based on the level of uncertainty in the prediction. This allows for a range of applications, such as refining semantic segmentation, correcting misclassified data, and improving model calibration.

Enabling the dropout layers during inference allows the model to make slightly different predictions for the same input each time. The variance between these predictions can be leveraged both in classification tasks (e.g., image and signal classification) to improve model accuracy and in segmentation tasks to generate uncertainty maps. By sampling multiple predictions during inference, the model can capture some of the uncertainty in its outputs. This Monte Carlo sampling approach has two main benefits:

- For classification tasks, averaging the predictions of multiple dropout samples can improve model accuracy compared to a single prediction.
- For segmentation tasks, the variance in the segmentation masks generated from different dropout samples provides a natural measure of uncertainty for each pixel or region. This produces an uncertainty map that highlights areas where the model is less confident.

Here we will discuss the key papers that have used MC dropout to improve their AI-based frameworks. Gal and Ghahramani [101] first proposed using MC dropout during inference to approximate Bayesian prediction intervals for neural networks. They showed that averaging the predictions from multiple dropout samples leads to improved classification accuracy and calibration of uncertainty estimates.

Jahmunah et al. [156] quantified uncertainty in an ECG model using MC dropout. This study develops a DenseNet model for myocardial infarction diagnosis from ECG signals that can quantify predictive uncertainty. Predictive entropy is computed based on the model's predictive probabilities and used as an uncertainty measure to detect misclassifications caused by out-of-distribution data. The results show that i) the model's uncertainty sensitivity increases as noise decreases, indicating increased confidence in predictions; ii) the model achieves high uncertainty accuracy and precision when SNR values are high, indicating it is aware of what it knows. Overall, MC dropout likely enables the model to estimate its predictive uncertainty, which allows it to detect misclassifications and indicate a lack of confidence when appropriate. This uncertainty awareness improves the model's reliability. Combalia et al., [121] used MC dropout and test time

augmentation to estimate prediction uncertainty for a skin lesion classification model. MC dropout allows the model to estimate how uncertain it is when classifying individual samples. The results show that uncertainty metrics based on MC dropout can detect difficult samples that the model tends to misclassify; identify out-of-distribution samples that differ from the training data. By removing the most uncertain samples, classification accuracy improved, indicating the uncertainty metrics can detect error-prone samples. In short, Monte Carlo dropout enables uncertainty estimation, which helps detect samples that may confuse the model. The uncertainty metrics based on it can improve model reliability, though challenges remain in identifying certain types of outliers. This study demonstrates that uncertainty estimation techniques based on MC dropout can enhance the performance and reliability of deep learning models for skin lesion classification.

The study conducted by Roy et al. [112] introduces a Bayesian convolutional neural network for whole-brain segmentation of MRI brain scans. The model uses Monte Carlo dropout at test time to generate samples from the posterior distribution, which allows it to estimate uncertainty in its segmentations. The entropy over the MC samples produces a voxel-wise uncertainty map while the mean of the MC predictions generates the final segmentation. In summary, Monte Carlo dropout enables the model to estimate uncertainty in its segmentations, which helps detect poor-quality segmentations. The structure-wise uncertainty metrics provide a useful interpretation of this uncertainty at the level of individual brain structures, facilitating both quality control and reliable analyses of large datasets. Edupuganti et al. [141] aim to quantify uncertainty in deep learning-based MRI reconstruction methods. The authors develop a variational autoencoder (VAE) to probabilistically reconstruct undersampled MRI scans. Monte Carlo sampling from the VAE's posterior distribution generates pixel variance maps, which quantify the uncertainty in the reconstructions. The MC dropout enables the VAE to i) model the uncertainty inherent in undersampled data; ii) quantify that uncertainty through pixel variance maps. The authors conclude that quantifying and reducing uncertainty in deep learning-based MRI reconstruction can improve diagnostic accuracy. In summary, MC dropout allows the VAE to model and visualize the uncertainty in its reconstructions, providing insights to improve model performance and reliability.

Overall, the MC dropout technique provides a practical and flexible approach to uncertainty quantification in deep learning, which is why it is the most used Bayesian method in this field.

4.4. Application areas

Fig. 9 illustrates the most employed methods in machine learning and deep learning for analyzing different anatomical regions. For models analyzing the thoracic and nervous systems, Bayesian inference emerges as the most utilized technique for uncertainty management. In the case of the digestive system, Dempster-Shafer theory (DST) is the predominant technique, while Monte Carlo simulation is found to be most used for analyzing diverse dataset collections. Regarding deep learning, the Bayesian method remains the primary technique across all analyzed organ systems.

Fig. 10 presents a sunburst diagram highlighting the prevalent deep models for the four most extensively studied image types in Fig. 6. Our analysis reveals a notable trend in the utilization of uncertainty techniques, with CNN models being the most employed for studying brain images, followed by Bayesian-based deep models. A similar pattern emerges for eye images, where CNN models are predominantly favored, closely followed by Bayesian-based deep models. In the case of skin images, CNN models hold a widespread preference, while Bayesian-deep models are the preferred choice for analyzing chest images.

Integrating uncertainty measurement into deep learning frameworks in healthcare can provide several benefits. It can help improve the reliability and interpretability of the model's predictions, enable better decision-making by clinicians, and enhance patient safety by highlighting areas of uncertainty in the model's output.

Fig. 11 illustrates the use of uncertainty techniques in the healthcare domain, incorporating machine learning and deep learning methods from 2013 to 2023. The graph reveals a growing trend in studies examining uncertainty using both machine learning and deep learning approaches throughout the years. Notably, there has been a significant increase in the application of uncertainty techniques in healthcare, particularly in 2019 and 2020, possibly driven by the need to analyze and detect conditions related to COVID-19 complications. However, there has been a decline in the number of studies focusing on uncertainty

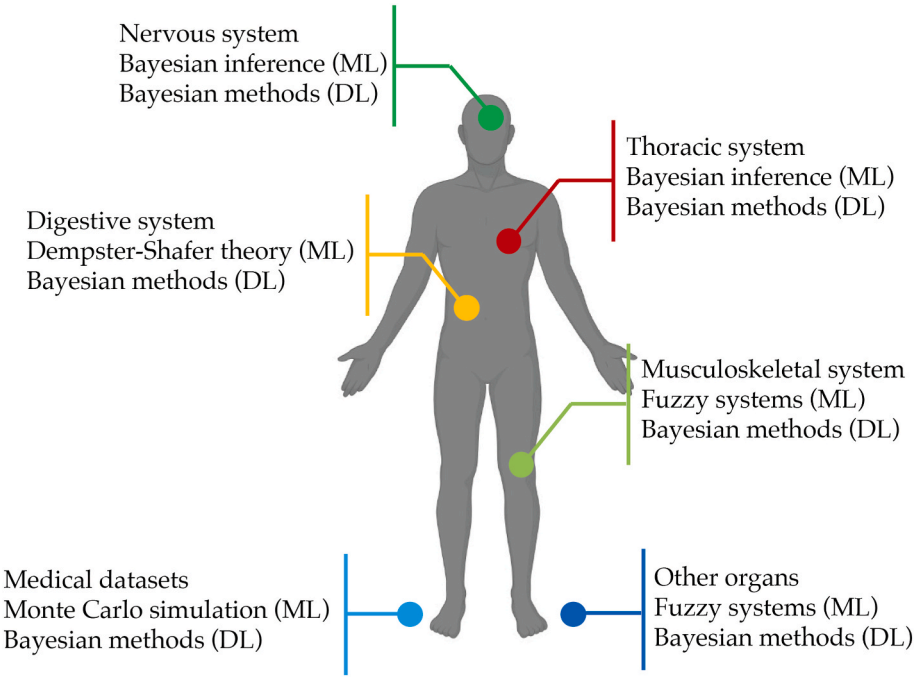

Fig. 9. Methods commonly employed for uncertainty estimation (both in machine learning and deep learning), categorized by different anatomical regions.

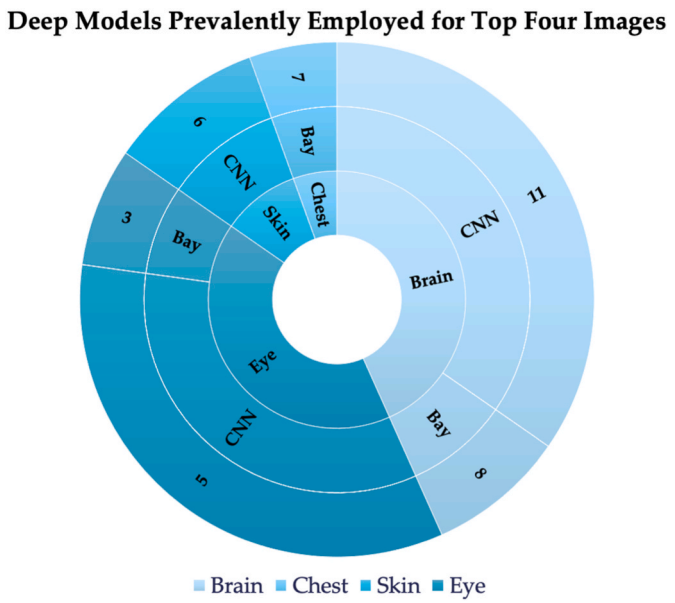

Fig. 10. Sunburst diagram detailing deep models most prevalently employed for the four top images studied in Table 3.
**The term 'Bay' refers to Bayesian-based deep models.

quantification in healthcare starting from 2021, which may be attributed to the rising adoption of uncertainty visualization techniques as reported in recent literature [208].

Based on the information presented in Table 2, it is evident that uncertainty techniques are commonly combined with machine learning approaches to capture and represent uncertainty in either data, models, or both. As a result, many studies focus on providing qualitative results, with only a few exploring methods for quantifying uncertainty. However, when examining deep learning approaches combined with uncertainty techniques, the emphasis is primarily on quantifying the inherent uncertainty in the data or the model, offering practical solutions for managing uncertainty in real-world medical systems. It is important to note that most studies investigate model uncertainties, followed by uncertainties in both the model and the data. The relatively fewer investigations into data uncertainties may be attributed to the fact that model uncertainties can be mitigated by improving the model architecture, learning process, and quality of training data, whereas data uncertainties are inherent and cannot be reduced [4]. As a result, researchers often prioritize refining their models to reduce uncertainties rather than exploring approaches to enhance training performance on noisy data. Finally, it should be emphasized that many authors who study model uncertainty employ the Monte Carlo dropout technique, which is computationally complex [209] and may pose limitations in healthcare settings where timely and rapid diagnoses are crucial.

This review study has some benefits and shortcomings, as discussed below:

4.4.1. *Advantages*

(i) This review summarizes recent research on uncertainty techniques in the machine and deep learning models in healthcare.
(ii) The type of diseases that have been studied using machine learning with uncertainty techniques have also been discussed.
(iii) The frequency of machine learning methods used with uncertainty techniques in the past decade has been examined.
(iv) The most used medical images for uncertainty techniques involving deep learning models in the past decade have been identified.
(v) The most used deep learning with uncertainty techniques for the top four studied images in the past decade have been identified.

4.4.2. *Limitation(s)*

(i) Uncertainty techniques used in healthcare involving animal or plant data were not considered in this review.

5. Future work

Based on the findings of the review, it is evident that further research is needed to explore uncertainty techniques in deep models for healthcare applications, particularly in relation to physiological signals. Estimating uncertainty is crucial for quantifying and effectively managing the inherent noise, interference, and imperfections present in 1D physiological data. This, in turn, improves the quality of measurements, resulting in more accurate and reliable outcomes. Additionally, uncertainty quantification has the potential to enhance the reliability of model predictions, even in scenarios involving missing or noisy data.

The existing studies have mainly focused on datasets of one type or very few multimodal data. Therefore, future investigations should delve into uncertainty techniques for multimodal data. In multimodal data involving diverse sources like images, text, and physiological signals, uncertainty can arise from various factors, including sensor quality, measurement accuracy, and inherent variability across modalities. By employing uncertainty techniques, confidence levels can be quantified for outcomes derived from each distinct modality and the integrated multimodal dataset as a whole. This proactive approach contributes to the refinement of predictions, ensuring heightened precision and robustness in the results obtained.

Most of the current studies primarily concentrate on binary classification or segmentation problems. For this reason, it is recommended that future works incorporate an assessment of quantitative uncertainties in classification probabilities for multiclass data. Expanding the evaluation to include multiclass scenarios would offer a more comprehensive understanding of uncertainty in classification tasks.

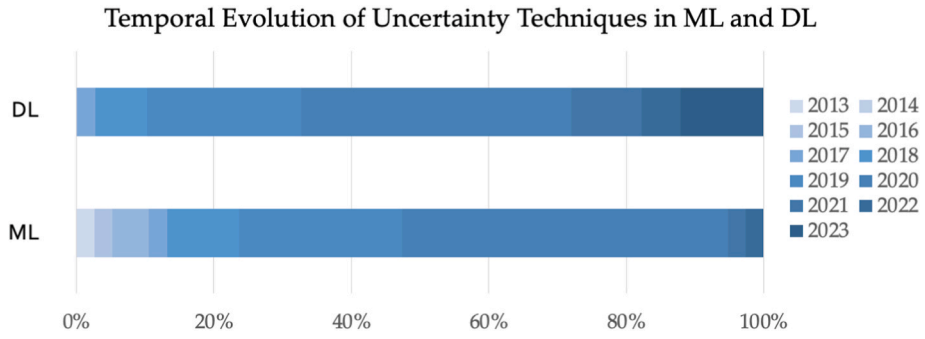

Fig. 11. Bar graphs of uncertainty techniques involving machines (top graph) and deep learning (bottom graph) from 2013 to 2023.
**The term 'DL' refers to deep learning models while 'ML' refers to machine learning models.

These avenues of research have the potential to enhance the accuracy and reliability of uncertainty techniques in healthcare applications.

To compare various uncertainty quantification methods and determine which one performs best on a given task, it is necessary to test them all on the same dataset. However, this review highlights a significant heterogeneity in both the tasks and datasets used across different studies. This variation can make it challenging to compare and draw general conclusions from the results. Furthermore, the use of different evaluation metrics and protocols across studies can further complicate comparisons. For example, some studies may report only accuracy, while others may report additional metrics such as precision, recall, or F1 score. Additionally, the choice of the dataset used for evaluation can significantly impact the results, as some datasets may be more challenging or have different characteristics than others.

To address these issues, future studies may benefit from using benchmark datasets and evaluation metrics to allow for more direct comparisons between different uncertainty quantification methods. Additionally, the development of challenges may help establish a standardized framework for evaluating the performance of these methods.

Future works could also investigate the following topics:

I. Development and exploration of UQ methods in AI models, especially in ML models where fewer studies exist: The field of machine learning offers several areas that warrant further exploration and research, including the development and exploration of UQ methods. Despite notable advancements in UQ for machine learning, there is still a need for more methods to be proposed and explored, especially in healthcare [210].
II. Fusion-based methods for enhancing AI techniques: Fusion-based methods, which combine multiple sources of information, offer a promising avenue for improving both predictions and uncertainty estimation in machine learning. Investigating and exploring fusion-based approaches further can provide insights into their potential benefits and applications [20], especially in healthcare [211,212].
III. Leveraging new theories for uncertainty quantification. The introduction of new theories can provide valuable frameworks for uncertainty quantification in machine and deep learning. For example, three-way decisions offer a decision-making approach that considers acceptance, rejection, and uncertainty as possible outcomes, making it a useful UQ method for tackling uncertain scenarios [213]. Similarly, info-gap decisions provide a theoretical foundation for decision-making in the face of severe uncertainty, where precise knowledge of the model or parameters may be lacking [214,215].
IV. Application of transfer learning techniques for uncertainty quantification. When data availability is limited, the application of transfer learning techniques becomes relevant for uncertainty quantification. Transfer learning enables leveraging knowledge and patterns acquired from a source domain with abundant data to enhance learning in a target domain with fewer samples. Investigating the effectiveness of transfer learning in the context of UQ can provide valuable insights and potential benefits.
V. Handling uncertainty in Graph Neural Networks (GNNs) and Graph CNNs. The advent of GNNs and Graph CNNs has introduced new challenges and opportunities in uncertainty quantification. These specialized architectures facilitate learning from graph-structured data, but efficient handling of uncertainty in such models requires the proposal and development of innovative methods specifically designed for GNNs [216] and Graph CNNs [217]. A review of existing techniques utilized in these domains can offer valuable insights and inform the development of novel approaches.
VI. Enhancing uncertainty calibration approaches in machine learning. Uncertainty calibration approaches play a pivotal role in machine learning by ensuring that predicted uncertainties align with empirical uncertainties, enabling dependable decision-making. Proposing novel uncertainty calibration methods can enhance the precision and utility of uncertainty estimates. A review of pertinent literature, including related review papers, can serve as a foundation for identifying and citing established calibration methods.
VII. Lastly, the accessibility of public data is essential for advancing machine learning research and promoting collaboration. Access to diverse and well-curated datasets enables researchers to benchmark and compare methods, ensuring reproducibility and fostering further progress in the field. As a result, efforts should be directed toward encouraging the release and sharing of public data, supporting initiatives such as open data platforms or collaborative data-sharing communities.

6. Conclusion

AI models are increasingly being utilized in healthcare, emphasizing the need to assess the reliability and safety of these systems. A crucial aspect of this assessment involves quantifying the uncertainty in the predictions made by AI models. This study systematically reviewed recent research that employed uncertainty techniques in healthcare applications of machine and deep learning, adhering to PRISMA guidelines.

This review identified Bayesian methods as the primary uncertainty techniques used in healthcare. Moreover, UQ techniques were more prevalent in healthcare applications using deep learning models compared to traditional machine learning models. These findings provide valuable insights for advancing UQ research in healthcare, and improving the reliability and safety of AI systems in this critical field.

Quantifying uncertainty in clinical AI implementation offers several advantages, including improved model accuracy by reducing misclassifications, identification of uncertain cases, enhanced model reliability and safety, and increased confidence among clinical operators, leading to greater acceptance and usage. The results of this study pave the way for future investigations in uncertainty quantification, strengthening the reliability and safety of AI systems in healthcare. Future studies could explore the examined UQ techniques in 1D physiological signals, encompassing multiclass or multimodal data, to further enhance UQ implementation. Additionally, comparing different uncertainty quantification techniques using standardized datasets and consistent metrics would enable a comprehensive analysis of these methods.

It is worth noting that this review focused solely on uncertainty techniques applied to healthcare data and did not include uncertainty techniques in animal or plant data or non-healthcare-specific applications. Future analyses could be conducted to incorporate these aspects and provide a more comprehensive understanding of uncertainty techniques across various domains.

Data availability

Not admissible.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Table A1
Summary of studies on the introduction of uncertainty techniques in healthcare applications using machine learning approaches.

Author, year	Features and methods	Data information	Findings/Results (%)
Giustinelli et al., 2022 [64]	• Imprecise/precise probability	Data collected from specific subsets of HRS respondents using so-called Experimental Modules	Imprecise/precise probability of successfully studying late-onset dementia
Ghesu et al. [49], 2021	• Dempster-Shafer theory	Chest X-ray images, the view-classification of abdominal ultrasound images, brain MR scans	Assessment of the image quality that combines the uncertainty measurement with probabilistic classification
Lin et al., 2020 [27]	• Bayesian inference • Markov chain Monte Carlo (MCMC) simulation	National food consumption database (nonylphenols residual in food, nonylphenols toxicity data)	Construction of a probabilistic risk assessment framework for dietary exposure to NP using Bayesian inference is a successful approach to showing effects on renal disease
Aakoyun et al., 2020 [29]	• Bayesian inference • Markov Chain Monte Carlo (MCMC) samplers	106 TC scans	Bayesian inference method is successful in predicting the maximum aneurysm diameter
Castellazi et al., 2020 [218]	• Adaptive Neuro Fuzzy Inference System • Unimodal and multimodal magnetic resonance features	77 MRI acquisitions: 33 patients with AD and 27 with VD	Unimodal and multimodal features are combined in the ML model for the classification of Alzheimer's disease and vascular dementia, achieving a prediction accuracy of 77.33%.
Das et al. [41] 2020	• Linguistic Neuro-Fuzzy with Feature Extraction (LNF-FE) • Features extraction with principal component analysis	Medical dataset: Pima Indian Diabetes (PID), Mammographic Mass, Breast Cancer, Heart Statlog, Liver, Blood transfusion Services, Haberman Nepal Breast Cancer	The analysis of medical data with a Linguistic Neuro-Fuzzy with Feature Extraction (LNF-FE) classifies diseases successfully
Vidhya et al., 2020 [42]	• Modified adaptive neuro-fuzzy inference system (M-ANFIS) • Entropy	Big Healthcare Data (Patient portals, research studies, electronic health records, wearable devices, etc)	The proposed technique performs better than other machine learning techniques
Kaur et al. [43] 2020	• Adaptive neuro-fuzzy inference system • Neuro-Fuzzy system	Medical data: osteoarthritis (OA) rheumatoid arthritis (RA) and osteonecrosis (ON) diseases.	Proposed system outperforms the fuzzy system in areas such as accuracy, sensitivity, and specificity
Sood et al. [219] 2020	• Linear Discriminant Analysis-Adaptive Neuro-Fuzzy Inference System (LDA-ANFIS)	The dengue-related data (10 attributes, such as fever, pain behind eyes and so on) and the heart data (6 parameters, such as ECG, HDL, sex)	The proposed method demonstrates efficient performance and uses several experimental and statistical methods
Tsai et al. [220] 2020	• Monte Carlo simulation • Computation of DNA damages caused by radiation	Simulation of radical's diffusion and reaction in chemical state and a multi-scale DNA model	GPU-based microscopic Monte Carlo simulation tool for DNA disease is advantageous
Salgado et al. [221], 2020	• Monte Carlo simulations	Local demographic and consumption data (from CESCAS I study)	Study of the impact of a lessening in sugar-sweetened beverage consumption on cardiovascular disease and diabetes
Lee et al. [36], 2020	• Monte Carlo Simulation using Geant4 Application	Images acquired on two breast phantoms with our Si/CZT Compton camera imaging system	Proposed method confirmed the feasibility of using a Compton camera for the detection of breast cancer
Shih et al. [37], 2020	• Monte Carlo simulation	Blood irradiator simulated using Monte Carlo simulation and MAGAT gel dosimeter	The proposed method can be used to ensure blood products can achieve an accurate delivery dosage.
Gasparini et al. [222], 2020	• Monte Carlo simulation	49 care practices of adults with chronic kidney disease	Proposed framework using Monte-Carlo simulation helps in the visitation process with respect to the health care utilization analysis
Zouh et al. [28], 2019	• Markov chain Monte Carlo algorithm • Bayesian framework • Monte Carlo algorithm	PET image with synthetic data.	Construction of a Bayesian framework for quantifying uncertainty in image reconstruction with Poisson data
Lee et al. [223], 2020	• TOPAS Monte Carlo simulation for monitoring of proton beam range	Energy spectra by the CZT camera	Monte Carlo simulation for proton beam range verification in cancer therapy
Achariya et al. [59], 2020	• Cuckoo search for the features selection • Rough set for the definition of the rule	Electrical information system of 606 patients with heart disease	Cuckoo search and rough set (CRCS) model developed for knowledge inference from information medical systems aided in the detection of heart disease.
Santra et al. [60], 2020	• Rough set-based lattice structure • Generation of the best set of decision rules during inference	Low back pain data	Proposed method can be extended to complex medical datasets for the representation of knowledge
Bania et al. [61] 2020	• R-Ensemble method based on rough set theory	Medical datasets, collected from UCI Machine repositories	Results demonstrate the superiority of the R-ensemble method over other attribute selection algorithms
Buono et al. [47], 2020	• Dempster-Shafer method for the diagnosis of skin diseases	Data on skin disease	The proposed method is successful in the diagnosis of skin diseases
Biswas et al. [54] 2020	• Soft fuzzy set • Dempster-Shafer method	600 chest x-ray image data set	Proposed method enhances the X-ray images of lungs and improves the visual quality of normal/diseased structural regions in X-ray images effectively.
Magnusson et al., 2019 [224]	• Bayesian inference • Bayesian inference through Markov chain Monte Carlo (MCMC) methods using the No-U-Turn sampler (NUTS).	1651 patients (administration of 2 mg siponimod or placebo)	Main stratum estimator based on Bayesian inference can be used to quantify the effect of treatment on disability progression in the (latent) population of patients with multiple sclerosis
Lipkova et al., 2019 [31]	• Bayesian framework	3D slice of the synthetic data (FET-PET image), and Clinical data from 8 patients with glioblastoma (GBM)	The proposed multimodal Bayesian model calibration is promising in assisting the development of personalized radiotherapy procedures
Flügge et al., 2019 [32]	• Bayesian networks • Approximate inference with variational message passing, loopy belief propagation, expectation propagation	Real diagnosis and anamnesis data	Bayesian network yields accurate headache diagnosis, with expectation propagation outperforming variational message passing.
Wang et al., 2020 [33] 2019	• Conditional Gaussian Bayesian network	15 variables from lung cancer patients (identified from 1996 to 2010)	Bayesian uncertainty estimation for lung cancer patient medical expenditure outperformed other investigated models.

(*continued on next page*)

Table A1 (*continued*)

Author, year	Features and methods	Data information	Findings/Results (%)
Liu et al., 2019 [225]	• Fuzzy inference logic model	Medical data associated with prostate cancers (1 933 535 items of structured and recognizable medical information from 8000 patients).	A fuzzy inference-based medical decision-making model aids reliable diagnosis.
Prameswari et al. [48], 2019	• Dempster-Shafer's theory • E-diagnostics for digestive system disorders	Data on the digestive diseases collected by experts	Web-based E-diagnostic for Digestive System Disorders using the Dempster Shafer's Method proved to yield higher certainty and accuracy (85%) in matching an expert's diagnosis
Razi et al. [226] 2019	• Dempster-Shafer theory Linear discriminant analysis (LDA)	EEG and EOG signals for 9 subjects	Results obtained show considerable improvement, highlighting the success of the proposed approach in modeling uncertainty, in multi-class classifications.
Porebski et al. [55], 2019	• Fuzzy focal elements • Dempster-Shafer theory	Medical database of patients affected by the hepatitis C virus	Proposed method provides simple diagnostic rules, helpful in processing inadequate data.
Xiao et al. [56] 2018	• The belief entropy • Fuzzy preference relations analysis • Dempster-Shafer theory	Medical diagnosis	Proposed method outperforms other related methods as uncertainty arising from human cognition can be lowered.
Shi et al. [50], 2018	• Dempster-Shafer theory of evidence	Drug-drug interaction dataset: 569 drugs and 52416 pairwise interactions between them	Combined Dempster-Shafer-based local classification models outperform single models for drug-drug interactions.
Kang et al. [51] 2018	• Gaussian mixture model (GMM) model • Dempster-Shafer theory	Dataset related to Clostridium difficile infection from 22 hospitals (Interior Health Authority (IHA), British Columbia)	Proposed model enables the generation of criteria ratings of risk factors to avert the imprecision caused by experts' judgments.
Mckenna et al. [65], 2018	• Uncertainty estimation with imprecise/precise probability	Breast cancer images	Proposed method proved that mathematical modeling for the optimization of chemotherapy for breast cancer therapy.
De Medeiros et al. [45], 2017	• Fuzzy inference system	Data from the World Bank, e.g Body mass index, Blood pressure, Physical activities, Eating habits	Proposed fuzzy logic supports real-time medical diagnosis and quantifies management.
Wang et al. [52], 2016	• Dempster-Shafer theory of evidence	Three datasets about decision-making in medical diagnosis	Proposed approach can reduce uncertainty caused by humans' subjective cognition
Mahmoud et al. [66], 2016	• Ensemble models • Creedal decision trees (CDTs) based on imprecise probability	Thrombosis Disease Dataset Hypothyroid Disease Database Arrhythmia Disease Database Heart Disease Database	Ensemble models yielded higher classification accuracy as compared to single-tree models.
Nguyen et al. [46], 2015	• Interval type-2 fuzzy logic system • Adaptive neuro-fuzzy inference system	The breast cancer database (699 breast cancers cases) The heart disease dataset (303 cases)	Wavelet transform and Interval type-2 fuzzy logic system successful for medical data classification.
Li et al. [58], 2015	• Fuzzy soft set • Dempster-Shafer theory of evidence	Dataset for decision-making in medical diagnosis problems	Fuzzy soft set and Dempster-Shafer theory practical for medical diagnosis.
Ghasemi et al. [57], 2013	• Fuzzy inference system • Dempster-Shafer theory	Simulated and real Brain MR images	Fuzzy inference and Dempster-Shafer for brain MRI segmentation yielded satisfactory results.

Table A2
Summary of studies on uncertainty estimation techniques in healthcare applications using deep learning approaches.

Author, year	Features and methods	Data information	Findings/Results (%)
Leibig et al., [105] 2017	• Bayesian deep neural network • Ensemble model • Monte Carlo dropout	Kaggle data: 35126 training and 53576 test fundus images; Messidor data: 1200 fundus images	Bayesian uncertainty measures determine uncertainty better than other direct methods.
Ktena et al., [89] 2017	• Convolutional neural network • Single deterministic methods	MRI images from 871 subjects	Proposed method improved overall classification by 11.9%
Ozdemir et al., [106] 2017	• Bayesian convolutional neural network • Bayesian convolutional neural network with uncertainty fusion	CT scan images of 888 patients' pulmonary nodules	Infusing uncertainty measures in the workflow improves prediction accuracy and model confidence.
Heo et al., [110] 2018	• Bayesian methods • Unique variational attention model	Physionet dataset: 36 physiological signals Pancreatic cancer dataset: 3699 patient records representing qualitative data Sepsis dataset: 22395 patient records comprising 14 variables	Proposed model yields large improvements compared to existing attention models.
Ayhan et al., [192] 2018	• Test-time augmentation methods	35126 training and 53576 test images	Proposed method provides useful approximations for the predictive uncertainties of deep models.
Wang et al., [191] 2018	• Test-time augmentation models	MRI images of 18 fetal patients, MRI images from 198 brain tumor patients	Proposed fine-tuning technique increases the segmentation accuracy and the method used in the study reduces user time and interactions.
McClure et al., [177] 2018	• Ensemble model	sMRI images from 5 datasets: 956, 1136, 183, 120 and 893 images from respective datasets	Distributed weight consolidation measure improves the performance of each independent test, as compared to the standard ensemble model.
Jungo et al., [107] 2018	• Monte Carlo dropout models	Brain tumor images of 46 subjects	Uncertainty information is obtainable by applying the Monte Carlo dropout after each convolutional layer.
Jungo et al., [108] 2018	• Bayesian methods • Monte Carlo dropout	30 MRI images of brain tumor	Learned observers' uncertainty can be fused with a Monte Carlo-based Bayesian network to determine the uncertainty of the model's parameters.

(*continued on next page*)

Table A2 (*continued*)

Author, year	Features and methods	Data information	Findings/Results (%)
Devries et al., [91] 2018	• Monte-Carlo dropout	2750 dermoscopic images	Heteroscedastic neural networks had the least improvement, while other uncertainty estimation techniques had similar results.
Dhamala et al. [161],2018	• Monte Carlo dropout	Body surface ECG and epicardial potentials	Quantifying the Uncertainty in Model Parameters Using Gaussian Process-Based Markov Chain Monte Carlo in Cardiac Electrophysiology
Tanno et al. [158], 2019	• Bayesian inference	Lifespan dataset, Prisma dataset, Human connection project dataset	Image enhancement based on uncertainty quantification in MRI data with a brain tumor (glioma) and multiple sclerosis
Graham et al. [195], 2019	• Test-time dropout	Gland Segmentation (GlaS) challenge dataset (MICAI) and a indipendent adenocarcinoma dataset	Segmentation in colon histology images
Jungo et al., [176] 2019	• Ensemble models • 5 uncertainty measures	Brain tumor images from 265 patients, skin lesion images dataset	While existing uncertainty measures calibrate well at the data level, they are not well calibrated at the subject level.
Orlando et al., [109] 2019	• Bayesian neural network with Monte Carlo dropout	OCT scans of 50 patients	Proposed uncertainty estimates highlight areas for model correction based on inverse performance correlation.
Wu et al., [178] 2019	• Ensemble methods	1660 magnetic resonance images of Alzheimer's disease	The proposed model is useful for uncertainty prediction for multi-class classification.
Adrian et al., [111] 2019	• Monte Carlo dropout	MRI images of 465 patients	Monte Carlo sample dropout identifies scans for further examination based on model errors.
Wang et al., [193] 2019	• Test-time augmentation • Test-time dropout	MRI scan images of 60 features	Aleatoric uncertainty estimation outperforms test-time dropout, reducing overconfident predictions.
Roy et al., [112] 2019	• Uncertainty map from Monte Carlo samples	4 datasets of brain images	The proposed uncertainty metrics could evaluate the accuracy of segmentation methods in deep models.
Ghesu et al., [93] 2019	• Bootstrapping for uncertainty • Single deterministic models	112120 and 185421 images from 2 datasets respectively	Proposed uncertainty method can eliminate training data, increasing the robustness and accuracy of the model and determining reader errors.
Baumgartner et al., [114] 2019	• Bayesian models	1018 thoracic CT images, prostate MR images from 68 patients	Proposed technique is able to produce a more naturalistic and varied segmentation of images as compared to other related works.
Raczkowski et al., [115] 2019	• Variational dropout-based entropy measure	5000 images of colorectal cancer	Variational dropout entropy measures increase the model's speed by about 45%.
Eaton-Rosen et al., [116] 2019	• Monte Carlo, varying thresholds of output confidence maps of a model, M-head uncertainty measures • Bayesian models	White-matter hyperintensity image data of 60 subjects	The proposed technique is effective in the counting of histopathological cells and white matter hyperintensity.
Jena et al., [118] 2019	• Bayesian neural network with Monte Carlo dropout	3 datasets of brain tumor, cell membrane and chest radiograph images	Proposed method enhances segmentation quality and uncertainty estimation accuracy compared to existing methods.
Soberanis-Mukul et al., [119] 2019	• Monte-Carlo dropout	Pancreas, spleen image datasets	The proposed method improves dice scores for both images compared to the original prediction by the model.
Hu et al., [120] 2019	• Variational dropout	1018 lung CT images, 48 prostate MRI images	Predictive uncertainty estimates, sample accuracy and diversity are improved.
Zhang et al., [194] 2019	• Test-time augmentation methods	11049 training and 5048 test MRI images of the knee	Proposed method reduces the reconstruction uncertainty of MRI images.
Sedghi et al., [149] 2019	• Bayesian methods	MRI brain images of 115 subjects	Intra-subject dice scores for grey matter, white matter and cerebrospinal fluid obtained were 0.70, 0.77 and 0.62 respectively.
Cortes-Cirianco et al., [201] 2019	• Ensembles of 100 deep neural network models • Test-time augmentation methods	2.035.207 bioactivity data points per protein	A strong relationship exists between confidence levels and the percentage of confidence intervals reflecting true bioactivity.
Cortes-Cirianco et al., [202] 2019	• Test-time dropout	4.795.207 bioactivity data points per protein for 24 target proteins	A strong relationship between confidence levels and error rates.
Norouzi et al., [150] 2019	• Monte-Carlo sampling • Computation of model uncertainty by estimation of the variance of segmentation results	7980 MRI images	Results demonstrate the successful integration of simple ideas with deep neural networks, highlighting their potential.
Filos et al., [151] 2019	• Bayesian deep learning techniques, convolutional neural network • Uncertainty estimation methods	35126 training images, 53576 test images	Comparing Bayesian deep learning techniques enables new methods to showcase efficacy on large-scale problems.
Tardy et al., [98] 2019	• Single deterministic method	In-house database: 1600 mammographies	Proposed method allows the rejection of the most obvious outliers and improved area under the curve results by up to 10%.
Jensen et al., [99] 2019	• Single deterministic method	31017 skin images	Proposed method provides improvements in model calibration, aiding in capturing uncertainty in samples and labels
Ghoshal et al., 2020 [152]	• Dropweights-based Bayesian Convolutional Neural Networks • Monte Carlo Dropweights (Bayesian convolutional neural networks	5941 Postero-Anterior chest radiography images	Uncertainty in prediction has a strong correlation with classification accuracy
Athanasiadis et al., [195] 2020	• Test-time augmentation methods	Audio-visual emotion datasets: 7386 audio recordings, 7442 videos and 96 images from 187 participants.	The best model achieved 52.52% classification accuracy in one dataset and 47.11% in another.

(*continued on next page*)

Table A2 (*continued*)

Author, year	Features and methods	Data information	Findings/Results (%)
Ayhan et al., [196] 2020	• Test-time augmentation methods • Computation of variance using entropy as a distribution of predicted probabilities	89215 fundus images	The proposed model achieved a high accuracy between 95.9 and 98.2%.
Graham et al., [94] 2020	• Single Deterministic methods • Measurement of cross-entropy uncertainty	593 MRI images	Obtained a dice score of about 0.85 for all regions in the uncertainty-aware hierarchical model
Hu et al., [124] 2020	• Zone-based uncertainty estimates based on Monte Carlo dropout technique	Scanned images from 83 patients	Obtained sensitivity of 74.7%
Nair et al., [125] 2020	• Computation of approximate probability distributions with Monte Carlo dropout	MRI images from 1064 patients	Overall lesion true-positive rate was at 0.8 and false detection rate was at 0.2
Natekar et al., [148] 2020	• Bayesian models	285 training cases, 48 testing volumes	Whole tumour dice coefficient obtained is 0.830
Wang et al., [144] 2020	• Bayesian models • Uncertainty as the mean of probabilistic predictions	5028 images	An accuracy of about 95% was obtained.
Scalia et al., [199] 2020	• Test-time augmentation methods • Monte Carlo dropout, deep ensembles, bootstrapping methods for quantification of prediction uncertainty	4 datasets of molecular graphs	Test set errors of 0.74, 0.32, 1.33 and 0.481 for 4 datasets respectively.
Sieradzki et al., [147] 2020	• Dropout-based uncertainty estimation • Derivation of uncertainty measured from variance in dropout	A deep generative model with recurrent neural networks	Proposed method enabled models to gain precision values between 0.0004 and 0.0007.
Laves et al., [123] 2020	• Deep Bayesian models • Variational inference with Monte-Carlo dropout	Variety of medical image datasets	Well-calibrated uncertainty in regression enables strong rejection of unreliable predictions or identification of out-of-distribution samples
Herzog et al., [113] 2020	• Bayesian convolutional neural network • Bayesian uncertainty	Magnetic resonance images of 511 patients with ischemic stroke	Bayesian methods improved image prediction, uncertainty estimation, and patient-level identification of uncertain aggregations. Bayesian network achieved 95% classification accuracy.
Luo et al., [92] 2020	• Single deterministic models	4 cardiac magnetic resonance image datasets	The recommended method yields the best quantification accuracy and optimization results.
Kwon et al., [126] 2020	• Bayesian neural network • Decomposition of predictive uncertainty into data and model uncertainty	Ischemic stroke lesion, retinal image datasets	Proposed uncertainty quantification technique provides deeper understanding of the point predictions
Selvan et al., [127] 2020	• Bayesian models	1018 thoracic CT images, 68 CT images	Proposed model captured richer segmentation variations, improving the quality and diversity of samples acquired
Hoebal et al., [181] 2020	• Ensemble models	CT images of lung nodules	Useful uncertainty information can be obtained when the model is trained using weighted categorical cross entropy measure.
Seebock et al., [128] 2020	• Bayesian U-net model • Monte Carlo dropout for epistemic uncertainty estimation	Six datasets of macula-centered spectralis	Images were segmented with high accuracy for healthy and diseased retinal images.
Hiasa et al., [129] 2020	• Bayesian U-net CNN model • Monte-Carlo dropout	20 fully annotated and 18 partially annotated CT images of hip and thigh	High uncertainty pixels relate to segmentation failure, allowing patient-specific muscle analysis.
Liao et al., [95] 2020	• Aleatoric uncertainty • Single deterministic methods	Images from 3157 patients	Absolute error of the model is reduced as compared to the conventional regression model.
Xia et al., [130] 2020	• Bayesian deep model	Pancreas and liver tumour data	Multi-view co-training on 2d models produces promising results.
Marc et al., [131] 2020	• Bayesian models	1018 lung CT scan images, MR prostate images from 68 patients	Proposed method requires lesser memory compared to not using reversible blocks, despite comparable segmentation accuracy being obtained.
Mehrtash et al., [182] 2020	• Model ensembling for calibration of model confidence	Brain, heart, and prostate MRI images	Model ensembling is successful in the confidence calibration of fully convolutional neural networks trained with dice loss.
Wickstrom et al., [132] 2020	• Monte-Carlo guided backpropagation	912 RGB colonoscopy images of 36 patients	Proposed method models input feature uncertainty, highlighting its contrast in correct and incorrect predictions.
Carneiro et al., [133] 2020	• Bayesian learning and inference, non-Bayesian inference • Uncertainty (entropy) and confidence calibration	940 colorectal polyps images	Proposed method yields good results pertaining to confidence calibration and classification accuracy.
Li et al., [134] 2020	• 3 Monte-Carlo dropout methods • Negative log likelihood, expected calibration error, Brier score	37 benign, 48 malignant images of the colon, 900 training and 379 testing skin images	Proposed region acquisition method improves model calibration compared to full region acquisition, irrespective of uncertainty measure.
Dahal et al., [183] 2020	• Ensemble model • Monte-Carlo, Horizontal stacked ensemble, test time augmentation • variance, entropy, mutual information, probabilistic atlas	2 datasets; cardiac ultrasound images from 500 patients and images taken from 10 030 varying echocardiography videos.	Uncertainty estimation has been proven to automatically reject images of poor quality and enhance segmentation results.
Li et al., [96] 2020	• DistDeepSHAP uncertainty assessment method for feature importance • Single deterministic models	Autism brain images from 4 datasets; 106, 175, 72, and 71 images respectively	Proposed method has the potential to determine biomarkers related to disease in neuroimaging data.
Quan et al., [135] 2020	• Bayesian uncertainty estimates • Ensemble of semi-supervised learning to correct noisy labels	Upper gastrointestinal images	Proposed method effectively enhances the recognition accuracy for authentic and noisy clinical data.

(*continued on next page*)

Table A2 (*continued*)

Author, year	Features and methods	Data information	Findings/Results (%)
Chiou et al., [184] 2020	• Ensemble models	MRI prostate lesion images from 60 patients suspected of having cancer, diffusion-weighted MRI images from 80 patients	Better image representations are obtained in segmentation for cancer characterization.
Wang et al., [136] 2020	• Teacher-student Bayesian deep model • Monte-Carlo dropout	100 MRI images of the left atrium, 210 CT scan images of the kidney	Proposed method outperforms existing semi-supervised uncertainty estimates on both datasets.
Ye et al., [97] 2020	• Single deterministic methods • Lasso bootstrap approach for uncertainty estimation	Brain diffusion magnetic resonance images from 25 subjects	Uncertainty measures relate to estimation errors and generate reasonable confidence intervals.
Bian et al., [137] 2020	• Bayesian models • Uncertainty aware cross-entropy loss, uncertainty aware self-training approach, uncertainty feature calibration method	OCT images from 623 to 537 patients respectively	Optimum results were obtained compared to existing methods for unsupervised domain adaptation tasks.
Araujo et al., [197] 2020	• Test-time augmentation	About 93 000 retinal images	Best result was obtained by the quadratic-weighted Cohen's kappa (between the range of 0.71–0.84).
Combalia et al., [121] 2020	• Monte-Carlo test-time dropout uncertainty estimation (epistemic and aleatoric) method	ISIC 2018 Challenge dataset: 10 015 dermoscopic images, 7470 skin lesions; ISIC 2019 Challenge dataset: 25331 training images, 8238 test images	Results demonstrate the successful use of uncertainty metrics for the detection of difficult and out-of-distribution samples.
Linmans et al., [179] 2020	• Monte- Carlo dropout • Deep ensemble models	26 whole slide images of breast cancer	Multi-head CNN outperforms MC dropout and deep ensembles, while the meta-loss function enhances out-of-distribution detection.
Toledo-Cortes et al., [122] 2020	• Bayesian models	EyePACS dataset	Proposed model yielded better results than the original deep learning model and enables uncertainty analysis.
Liang et al., [180] 2020	• Ensemble methods	4 publicly available datasets: head CT, mammography, chest x-ray, histological images	Proposed approach reduces calibration error largely across the different models and datasets
Stoean et al. [157], 2020	• Monte Carlo dropout	Eighty-five EOG tests	The novel method integrates uncertainty quantification into decision trees using MC dropout, achieving 81.18% accuracy in classifying control, presymptomatic, and sick classes
Guo et al. [158], 2020	• Monte Carlo dropout	The UKBB dataset: 3D cardiac images	Incorporating MCD uncertainty enhanced the segmentation performance of the model when applied to cardiovascular disease image data.
Guo et al. [188], 2020	• Ensemble models	Public CXR dataset (15134 images)	The method combines label fusion, uncertainty-guided continuous kernel cut, and deep learning for accurate ventricle segmentation and function measurements in cardiac cine MRI.
Corrado et al. [169], 2020	• Bayesian probability approach	Cardiac MRI	Quantifying atrial anatomy uncertainty from clinical data and its impact on electro-physiology simulation predictions
Tanno et al., [138] 2021	• Bayesian inference for uncertainty	288 diffusion-weighted MRI images of the brain per subject, MRI images of 26 subjects, 2 healthy male MRI images, brain tumor + multiple sclerosis images	Uncertainty measurement enhances prediction, detects failures, and provides insightful explanations for model performance.
Cao et al., [185] 2021	• Ensemble model	13382 ultrasound images from 107 patients	The proposed model obtained a high classification accuracy of 99.21%.
Ghoshal et al. [140], 2021	• Monte-Carlo DropWeights • Bayesian Residual UNet	Segmentation task: Dataset from Kaggle Data Science Bowl Challenge 2018 Classification task: 96115 MRI images of medical images	Monte-Carlo DropWeights and Bayesian Residual UNet for uncertainty estimation in medical image segmentation and classification.
Edupuganti et al., [141] 2021	• Monte-Carlo sampling technique	320 2d image slices per patient, from 19 patients	A high SURE-MSE value of 0.97 was achieved for 2-fold under sampling
Qin et al., [186] 2021	• Brain diffusion MRI images • Convolutional neural network • Measurement of variance in results combined from the training of ensemble deep models	About 1, 000, 000 images	Correlations between estimation uncertainty and error were considerable, $p < 0.001$.
Valliuddin et al., al., [142] 2021	• Bayesian models	1000 polyp images	Proposed approach increased predictive performance by up to 14%.
Teng et al., [143] 2021	• Bayesian method	Alzheimer's disease: 1574 patients Parkinson's disease: 1093 patients	Accuracy for Alzheimer's disease: 91.6% Accuracy for Parkinson's disease: 79.7%
Zhang et al., [145] 2021	• Bayesian method	A sample size of active class: 7039 A sample size of inactive class: 89922	Average area under the receiver operating characteristic curve: 0.734.
Vranken et al., [146] 2021	• Bayesian method	526656 ECG signals from three different datasets	Variational inference with Bayesian decomposition and ensemble outperforms other methods. High uncertainty in deep ECG classification correlates with lower diagnostic agreement.
Abdar et al., [198] 2021	• Monte-Carlo dropout, ensemble Monte-Carlo dropout, Deep ensemble uncertainty quantification techniques	Dataset 1: Kaggle skin dataset (2637 training, 660 test images). Dataset 2: ISIC dataset 2 (7234 training, 1808 test images)	The highest accuracy of about 91% was obtained in the second dataset.
Dong et al. [200], 2021	• Test-time augmentation	Public CXR dataset (15134)	The MUL method combines parallel dropout networks for accurate diagnoses and uncertainty estimations. RCoNet model outperforms existing methods in all metrics

(*continued on next page*)

Table A2 (*continued*)

Author, year	Features and methods	Data information	Findings/Results (%)
Thiagarajan et al., [139] 2022	• Bayesian-based, transfer learning CNN models	162 slide images of breast cancer	Bayesian CNN is advantageous over existing models and is useful in explaining the uncertainties in histological images.
Dolezal et al., al., [153] 2022	• Bayesian method	941 images	Confident predictions outperform uncertainty-free predictions. Thresholding reliably predicts lung adenocarcinoma and squamous cell carcinoma out-of-distribution data.
Mensah et al., [154] 2022	• Bayesian capsule network • Aleatoric and epistemic uncertainty estimation	Three computer vision datasets and one COVID-19 chest x-ray image dataset (16952 training, 2000 validation and 4227 test images)	Bayesian capsule networks have the potential to demonstrate the necessary transparency, credibility, reliability and interpretability to gain the confidence of industry partners.
Mazoure et al., [155] 2022	• Ensemble models	Skin lesion images from International Skin Imaging Collaboration (ISIC) 2018 dataset	Proposed method serves as the best-performing supervised and self-supervised and uncertainty estimation technique.
Singh et al., [187] 2022	• Ensemble models • Monte-Carlo dropout, test time augmentation techniques	Skin lesion images from International Skin Imaging Collaboration (ISIC) 2018 dataset (10 015 dermoscopic images)	The results demonstrate the robustness, transparency, and confidence of the proposed model.
Wang et al. [168], 2022	• Bayesian probability approach	200 CT images	A probabilistic generative approach for combining shape and intensity models for cochlear segmentation in CT images
Abdar et al. [161] 2023	• Monte-Carlo dropout	CT scan and X-ray images	*UncertaintyFuseNet* integrates EMCD for accurate classification of COVID-19 CT scan and X-ray images, achieving high accuracies of 99.08% and 96.35%.
Da Silvia et al. [159], 2023	• Monte-Carlo dropout	ECG data	The Monte Carlo method was used to identify the primary source of ECG measurement uncertainty, improving the understanding of the metrological behavior of ECG measurements.
Nasir et al. [160] 2023	• Monte-Carlo dropout	Real-world electronic health record (EHR) data	A model for early prediction of type 2 diabetes mellitus utilized real-world EHR data and incorporated Monte Carlo dropout for uncertainty estimation. The model showed a 1.6% accuracy improvement.
MacDonald et al. [162], 2023	• Monte-Carlo dropout	Transcriptomic data: three RNA-seq datasets	Three Bayesian DL models were compared for cancer prediction. Bayesian DL has the potential to improve performance, transparency, and safety by effectively handling uncertainty in real-world applications.
Farooq et al. [163], 2023	• Monte-Carlo	2 database BUSI and UDIAT datasets	A residual-attention-based uncertainty-guided mean teacher framework outperformed existing methods in breast ultrasound mass segmentation
Buddenkotte et al. [189], 2023	• Ensemble model	(i) CT images of high-grade serous ovarian cancer patients, (ii) a public dataset of CT images of kidney tumors	A scalable and intuitive framework for UQ in 3D medical image segmentation of cancers (ovarian and kidney)
Abdar et al. [166], 2023	• Bayesian model	Cardiac Syndrome X (CSX) dataset from Tehran's Heart Center	A binarized multi-gate mixture of Bayesian experts for cardiac syndrome X Diagnosis
Chen et al. [171], 2023	• Bayesian inference	Brain MRI (inter-patients 260 T1 and 451 T1 from two different datasets), and phantom images	An innovative transformer for unsupervised medical image registration
Zakeri et al. [165], 2023	• Bayesian inference	Cine cardiac magnetic resonance: UK Biobank (UKB) LAX cine CMR images	Learning-based deformable registration for realistic cardiac MR sequence generation from a single frame
Dolezal et al. [173], 2023	• Bayesian inference	Histologic images	Digital Histopathology with Real-Time Whole-Slide Visualization
Abdullah et al. [172], 2023	• Bayesian deep learning model • Monte Carlo Dropout out	Monte Carlo dropout	Histological images and ultrasound images of breast cancers
Abdar et al. [164], 2023	• Bayesian approximation	Retinal OCT, lung CT, and chest X-ray	Classification in Retinal OCT, lung CT, and chest X-ray
Jahmunah et al., [156] 2023	• Bayesian model	12 lead ECG signals from 148 MI patients and 52 healthy subjects (multiclass data)	The proposed model reliably presents diagnostic information, making it suitable for healthcare applications.

References

[1] I.H. Sarker, Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions, SN Comput. Sci. 2 (6) (2021) 420, https://doi.org/10.1007/s42979-021-00815-1.

[2] L. Wells, T. Bednarz, Explainable AI and reinforcement learning—a systematic review of current approaches and trends, Front. Artif. Intell. 4 (2021), https://doi.org/10.3389/frai.2021.550030.

[3] D. Seuß, Bridging the Gap between Explainable AI and Uncertainty Quantification to Enhance Trustability, 2021, pp. 1–10.

[4] J. Gawlikowski, et al., A Survey of Uncertainty in Deep Neural Networks, 2021, pp. 1–41.

[5] D. Amodei, C. Olah, J. Steinhardt, P.F. Christiano, J. Schulman, D. Mané, Concrete problems in AI safety, ArXiv abs/1606.0 (2016).

[6] M. Rußwurm, M. Ali, X.X. Zhu, Y. Gal, M. Körner, Model and data uncertainty for satellite time series forecasting with deep recurrent models, in: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, pp. 7025–7028, https://doi.org/10.1109/IGARSS39084.2020.9323890.

[7] Y. Gal, R. Islam, Z. Ghahramani, Deep Bayesian active learning with image data, in: 34th International Conference on Machine Learning, ICML 2017 vol. 3, 2017, pp. 1923–1932.

[8] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: representing model uncertainty in deep learning, in: 33rd International Conference on Machine Learning, ICML 2016 vol. 3, 2016, pp. 1651–1660.

[9] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter? Struct. Saf. 31 (2) (2009) 105–112, https://doi.org/10.1016/j.strusafe.2008.06.020.

[10] J. Mukhoti, Y. Gal, Evaluating Bayesian Deep Learning Methods for Semantic Segmentation, 2018.

[11] A. Malinin, M.J.F. Gales, Uncertainty Estimation in Deep Learning with Application to Spoken Language Assessment, May, 2019.

[12] A. Ashukha, A. Lyzhov, D. Molchanov, D. Vetrov, Pitfalls of In-Domain Uncertainty Estimation and Ensembling in Deep Learning, 2020, pp. 1–30.

[13] Y. Ovadia, et al., Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty under Dataset Shift, Jun. 2019, https://doi.org/10.48550/arxiv.1906.02530.

[14] D. Hendrycks and K. Gimpel, "A Baseline for Detecting Misclassified and Out-Of-Distribution Examples in Neural Networks.".

[15] S. Liang, Y. Li, R. Srikant, Enhancing the reliability of out-of-distribution image detection in neural networks, in: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018, pp. 1–15.

[16] H. Broekhuizen, C.G.M. Groothuis-Oudshoorn, J.A. van Til, J.M. Hummel, M. J. IJzerman, A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions, Pharmacoeconomics 33 (5) (May 28, 2015) 445–455, https://doi.org/10.1007/s40273-014-0251-x. Springer International Publishing.

[17] B. Lambert, F. Forbes, A. Tucholka, S. Doyle, H. Dehaene, M. Dojat, Trustworthy Clinical AI Solutions: a Unified Review of Uncertainty Quantification in Deep Learning Models for Medical Image Analysis, Oct. 2022.

[18] T.J. Loftus, et al., Uncertainty-aware deep learning in healthcare: a scoping review, PLOS Digit. Health 1 (8) (Aug. 2022), e0000085, https://doi.org/10.1371/journal.pdig.0000085.

[19] J. Gawlikowski, et al., A survey of uncertainty in deep neural networks [Online]. Available: http://arxiv.org/abs/2107.03342, Jul. 2021.

[20] M. Abdar, et al., A review of uncertainty quantification in deep learning: techniques, applications and challenges, Inf. Fusion 76 (Dec. 01, 2021) 243–297, https://doi.org/10.1016/j.inffus.2021.05.008. Elsevier B.V.

[21] E. Ocampo, M. Maceiras, S. Herrera, C. Maurente, D. Rodríguez, M.A. Sicilia, Comparing Bayesian inference and case-based reasoning as support techniques in the diagnosis of Acute Bacterial Meningitis, Expert Syst. Appl. 38 (8) (2011) 10343–10354, https://doi.org/10.1016/j.eswa.2011.02.055.

[22] Mooney, Z. Christopher, Monte Carlo Simulation, Sage, 1997.

[23] D. Karaboga, E. Kaya, Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey, Artif. Intell. Rev. 52 (4) (2019) 2263–2293, https://doi.org/10.1007/s10462-017-9610-2.

[24] T. Denœux, 40 years of Dempster–Shafer theory, Int. J. Approx. Reason. 79 (Dec. 2016) 1–6, https://doi.org/10.1016/J.IJAR.2016.07.010.

[25] B. Walczak, D.L. Massart, Rough sets theory, Chemometr. Intell. Lab. Syst. 47 (1999) 1999–2000.

[26] T. Augustin, F.P. Coolen, G. De Cooman, M.C. Troffaes, Introduction to Imprecise Probabilities, John Wiley & Sons, 2014.

[27] H.-C. Lin, et al., Bayesian inference of nonylphenol exposure for assessing human dietary risk, Sci. Total Environ. 713 (2020), 136710, https://doi.org/10.1016/j.scitotenv.2020.136710.

[28] Q. Zhou, T. Yu, X. Zhang, J. Li, Bayesian inference and uncertainty quantification for medical image reconstruction with Poisson data, SIAM J. Imag. Sci. 13 (1) (Jan. 2020) 29–52, https://doi.org/10.1137/19M1248352.

[29] E. Akkoyun, S.T. Kwon, A.C. Acar, W. Lee, S. Baek, Predicting abdominal aortic aneurysm growth using patient-oriented growth models with two-step Bayesian inference, Comput. Biol. Med. 117 (C) (Feb. 2020), https://doi.org/10.1016/j.compbiomed.2020.103620.

[30] B.P. Magnusson, H. Schmidli, N. Rouyrre, D.O. Scharfstein, Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by post-randomization events, Stat. Med. 38 (23) (2019) 4761–4771, https://doi.org/10.1002/sim.8333.

[31] J. Lipkova, et al., Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and Bayesian inference, IEEE Trans. Med. Imag. 38 (8) (2019) 1875–1884, https://doi.org/10.1109/TMI.2019.2902044.

[32] S. Flügge, S. Zimmer, U. Petersohn, Knowledge Representation and Diagnostic Inference Using Bayesian Networks in the Medical Discourse, 2019.

[33] K.-J. Wang, J.-L. Chen, K.-M. Wang, Medical expenditure estimation by Bayesian network for lung cancer patients at different severity stages, Comput. Biol. Med. 106 (Mar. 2019) 97–105, https://doi.org/10.1016/j.compbiomed.2019.01.015.

[34] M.V. Salgado, et al., Projected impact of a reduction in sugar-sweetened beverage consumption on diabetes and cardiovascular disease in Argentina: a modeling study, PLoS Med. 17 (7) (Jul. 2020), e1003224, https://doi.org/10.1371/JOURNAL.PMED.1003224.

[35] M.Y. Tsai, et al., A new open-source GPU-based microscopic Monte Carlo simulation tool for the calculations of DNA damages caused by ionizing radiation — Part I: core algorithm and validation, Med. Phys. 47 (4) (Apr. 2020) 1958–1970, https://doi.org/10.1002/MP.14037.

[36] Y. Lee, Preliminary evaluation of dual-head Compton camera with Si/CZT material for breast cancer detection: Monte Carlo simulation study, Optik 202 (Feb. 2020), https://doi.org/10.1016/J.IJLEO.2019.163519.

[37] T.Y. Shih, Y.L. Liu, H.H. Chen, J. Wu, Dose evaluation of a blood irradiator using Monte Carlo simulation and MAGAT gel dosimeter, Nucl. Instrum. Methods Phys. Res. 954 (Feb. 2020), https://doi.org/10.1016/J.NIMA.2018.09.084.

[38] A. Gasparini, et al., Mixed-effects models for health care longitudinal data with an informative visiting process: a Monte Carlo simulation study, Stat. Neerl. 74 (1) (Feb. 2020) 5–23, https://doi.org/10.1111/STAN.12188.

[39] C. Lee, K.P. Kim, D.J. Long, W.E. Bolch, Organ doses for reference pediatric and adolescent patients undergoing computed tomography estimated by Monte Carlo simulation, Med. Phys. 39 (4) (2012) 2129–2146, https://doi.org/10.1118/1.3693052.

[40] G. Castellazzi, et al., A machine learning approach for the differential diagnosis of alzheimer and vascular dementia fed by MRI selected features, Front. Neuroinf. 14 (Jun. 2020), https://doi.org/10.3389/fninf.2020.00025.

[41] H. Das, B. Naik, H.S. Behera, Medical disease analysis using neuro-fuzzy with feature extraction model for classification, Inform. Med. Unlocked 18 (Jan. 2020), https://doi.org/10.1016/j.imu.2019.100288.

[42] K. Vidhya, R. Shanmugalakshmi, Modified adaptive neuro-fuzzy inference system (M-ANFIS) based multi-disease analysis of healthcare Big Data, J. Supercomput. 76 (11) (Nov. 2020) 8657–8678, https://doi.org/10.1007/s11227-019-03132-w.

[43] R. Kaur, K. Kaur, A. Khamparia, D. Anand, An improved and adaptive approach in ANFIS to predict knee diseases, Int. J. Healthc. Inf. Syst. Inf. 15 (2) (Apr. 2020) 22–37, https://doi.org/10.4018/IJHISI.2020040102.

[44] K. Liu, et al., Big medical data decision-making intelligent system exploiting fuzzy inference logic for prostate cancer in developing countries, IEEE Access 7 (2019) 2348–2363, https://doi.org/10.1109/ACCESS.2018.2886198.

[45] I.B. de Medeiros, M.A.S. Machado, W.J. Damasceno, A.M. Caldeira, R. C. dos Santos, J.B. da Silva Filho, A fuzzy inference system to support medical diagnosis in real time, Procedia Comput. Sci. 122 (2017) 167–173, https://doi.org/10.1016/j.procs.2017.11.356.

[46] T. Nguyen, A. Khosravi, D. Creighton, S. Nahavandi, Medical data classification using interval type-2 fuzzy logic system and wavelets, Appl. Soft Comput. J. 30 (2015) 812–822, https://doi.org/10.1016/j.asoc.2015.02.016.

[47] M.L.C. Buono, N. Pandiangan, S.H.D. Loppies, The implementation of an expert system in diagnosing skin diseases using the dempster-shafer method, J. Phys. Conf. Ser. 1569 (Jul. 2020) 2, https://doi.org/10.1088/1742-6596/1569/2/022028.

[48] E.A. Prameswari, A. Triayudi, I.D. Sholihati, Web-based E-Diagnostic for Digestive System Disorders in Mumans Using the Demster Shafer Method, 2019.

[49] S. Razi, M.R. Karami Mollaei, J. Ghasemi, A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory, Inf. Sci. 484 (May 2019) 14–26, https://doi.org/10.1016/J.INS.2019.01.053.

[50] J.Y. Shi, X.Q. Shang, K. Gao, S.W. Zhang, S.M. Yiu, An integrated local classification model of predicting drug-drug interactions via dempster-shafer theory of evidence, Sci. Rep. 8 (Dec. 2018) 1, https://doi.org/10.1038/S41598-018-30189-Z.

[51] B. Kang, G. Chhipi-Shrestha, Y. Deng, J. Mori, K. Hewage, R. Sadiq, Development of a predictive model for Clostridium difficile infection incidence in hospitals using Gaussian mixture model and Dempster–Shafer theory, Stoch. Environ. Res. Risk Assess. 32 (6) (2018) 1743–1758, https://doi.org/10.1007/s00477-017-1459-z.

[52] J. Wang, Y. Hu, F. Xiao, X. Deng, Y. Deng, A novel method to use fuzzy soft sets in decision making based on ambiguity measure and Dempster–Shafer theory of evidence: an application in medical diagnosis, Artif. Intell. Med. 69 (May 2016) 1–11, https://doi.org/10.1016/J.ARTMED.2016.04.004.

[53] F.C. Ghesu, et al., Quantifying and leveraging predictive uncertainty for medical image assessment [Online]. Available: http://arxiv.org/abs/2007.04258, Jul. 2020.

[54] B. Biswas, S.K. Ghosh, S. Bhattacharyya, J. Platos, V. Snasel, A. Chakrabarti, Chest X-ray enhancement to interpret pneumonia malformation based on fuzzy soft set and Dempster–Shafer theory of evidence, Appl. Soft Comput. J. 86 (Jan. 2020), https://doi.org/10.1016/J.ASOC.2019.105889.

[55] S. Porebski, P. Porwik, E. Straszecka, T. Orczyk, Liver fibrosis diagnosis support using the Dempster–Shafer theory extended for fuzzy focal elements, Eng. Appl. Artif. Intell. 76 (Nov. 2018) 67–79, https://doi.org/10.1016/j.engappai.2018.09.004.

[56] F. Xiao, A hybrid fuzzy soft sets decision making method in medical diagnosis, IEEE Access 6 (2018) 25300–25312, https://doi.org/10.1109/ACCESS.2018.2820099.

[57] J. Ghasemi, R. Ghaderi, M.R.K. Mollaei, S.A. Hojjatoleslami, A novel fuzzy Dempster-Shafer inference system for brain MRI segmentation, undefined 223 (Feb. 2013) 205–220, https://doi.org/10.1016/J.INS.2012.08.026.

[58] Z. Li, G. Wen, N. Xie, An approach to fuzzy soft sets in decision making based on grey relational analysis and Dempster–Shafer theory of evidence: an application in medical diagnosis, Artif. Intell. Med. 64 (3) (Jul. 2015) 161–171, https://doi.org/10.1016/J.ARTMED.2015.05.002.

[59] K. A. P, D.P. Acharjya, A hybrid scheme for heart disease diagnosis using rough set and cuckoo search technique, J. Med. Syst. 44 (1) (Jan. 2020), https://doi.org/10.1007/s10916-019-1497-9.

[60] D. Santra, S.K. Basu, J.K. Mandal, S. Goswami, Rough set based lattice structure for knowledge representation in medical expert systems: low back pain management case study, Expert Syst. Appl. 145 (May 2020), https://doi.org/10.1016/J.ESWA.2019.113084.

[61] R.K. Bania, A. Halder, R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data,", Comput. Methods Progr. Biomed. 184 (Feb. 2020) https://doi.org/10.1016/J.CMPB.2019.105122.

[62] C. Blake, UCI Repository of Machine Learning Databases, 1998 ics. uci. edu/~mlearn/MLRepository. html.

[63] Q.Y. Jiang, X.J. Yang, X.S. Sun, An aided diagnosis model of sub-health based on rough set and fuzzy mathematics: a case of TCM, J. Intell. Fuzzy Syst. 32 (6) (2017) 4135–4143, https://doi.org/10.3233/JIFS-15958.

[64] P. Giustinelli, C.F. Manski, F. Molinari, Precise or imprecise probabilities? Evidence from survey response related to late-onset dementia, J. Eur. Econ. Assoc. 20 (1) (Feb. 2022) 187–221, https://doi.org/10.1093/jeea/jvab023.

[65] M.T. McKenna, J.A. Weis, A. Brock, V. Quaranta, T.E. Yankeelov, Precision medicine with imprecise therapy: computational modeling for chemotherapy in breast cancer, Transl. Oncol. 11 (3) (Jun. 2018) 732–742, https://doi.org/10.1016/J.TRANON.2018.03.009.

[66] A.M. Mahmoud, Suitability of various intelligent tree based classifiers for diagnosing noisy medical data, Egypt. Comput. Sci. J. 40 (2) (2016).
[67] A. Kendall, Y. Gal, What uncertainties do we need in Bayesian deep learning for computer vision?, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, in NIPS'17 Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 5580–5590.
[68] A. Malinin, M. Gales, Predictive uncertainty estimation via prior networks, Adv. Neural Inf. Process. Syst. 2018 (2018) 7047–7058. Decem, no. NeurIPS.
[69] M. Seçkin Ayhan and P. Berens, "Test-time Data Augmentation for Estimation of Heteroscedastic Aleatoric Uncertainty in Deep Neural Networks"..
[70] Neal, M. Radford, Bayesian learning for neural networks, Springer Sci. Bus. Media 118 (2012).
[71] J. Maroñas, R. Paredes, and D. Ramos, "Calibration of Deep Probabilistic Models with Decoupled Bayesian Neural Networks", doi: 10.1016/j.neucom.2020.04.103..
[72] C. Blundell, J. Cornebise, K. Kavukcuoglu, W. Com, and G. Deepmind, "Weight Uncertainty in Neural Networks Daan Wierstra"..
[73] B. Lakshminarayanan, A. Pritzel, and C. B. Deepmind, "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles"..
[74] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks, Neurocomputing 338 (2019) 34–45, https://doi.org/10.1016/j.neucom.2019.01.103.
[75] C. Guo, G. Pleiss, Y. Sun, K.Q. Weinberger, On calibration of modern neural networks, 34th Int. Conf. Mach. Learn. ICML 3 (2017) 2130–2143, 2017.
[76] K. Lee, H. Lee, K. Lee, J. Shin, Training confidence-calibrated classifiers for detecting out-of-distribution samples, 6th Int. Conf. Learn. Representations, ICLR 2018 Conf. Track Proc. (2018) 1–16.
[77] S. Thulasidasan, G. Chennupati, J. Bilmes, T. Bhattacharya, S. Michalak, On mixup training: improved calibration and predictive uncertainty for deep neural networks, Adv. Neural Inf. Process. Syst. 32 (2019) 1–15. NeurIPS.
[78] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826, https://doi.org/10.1109/CVPR.2016.308.
[79] B. Zadrozny, C. Elkan, Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers, Icml (March 2013) 1–8, 2001.
[80] J. Wenger, H. Kjellström, R. Triebel, Non-Parametric Calibration for Classification, 108, 2019.
[81] J. Zhang, B. Kailkhura, T. Yong-Jin Han, Mix-n-match: ensemble and compositional methods for uncertainty calibration in deep learning, in: 37th International Conference on Machine Learning, ICML 2020, 2020, pp. 11051–11062. PartF16814.
[82] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, Adv. Neural Inf. Process. Syst. 2017 (2017) 6403–6414. Decem, no. Nips.
[83] P. Izmailov, W.J. Maddox, P. Kirichenko, T. Garipov, D. Vetrov, A.G. Wilson, Subspace inference for Bayesian deep learning, in: 35th Conference on Uncertainty in Artificial Intelligence, UAI 2019, 2019.
[84] P. Oberdiek, M. Rottmann, H. Gottschalk, Classification uncertainty of deep neural networks based on gradient information, Lect. Notes Comput. Sci. 11081 LNAI (2018) 113–125, https://doi.org/10.1007/978-3-319-99978-4_9.
[85] J. Lee, G. Alregib, Gradients as a measure of uncertainty in neural networks, in: Proceedings - International Conference on Image Processing, ICIP, Vol. 2020-Octob, 2020, pp. 2416–2420, https://doi.org/10.1109/ICIP40778.2020.9190679.
[86] M. Raghu, et al., Direct uncertainty prediction for medical second opinions, in: 36th International Conference on Machine Learning, ICML 2019, 2019, pp. 9202–9211, 2019-June.
[87] T. Ramalho, M. Miranda, Density estimation in representation space to predict model uncertainty, Commun. Comput. Inf. Sci. 1272 (2020) 84–96, https://doi.org/10.1007/978-3-030-62144-5_7.
[88] M. Sensoy, L. Kaplan, M. Kandemir, Evidential deep learning to quantify classification uncertainty, Adv. Neural Inf. Process. Syst. (2018) 3179–3189, 2018-Decem, no. Nips.
[89] S.I. Ktena, et al., Distance metric learning using graph convolutional networks: application to functional brain networks, Lect. Notes Comput. Sci. 10433 LNCS (2017) 469–477, https://doi.org/10.1007/978-3-319-66182-7_54.
[90] R. McKinley, et al., Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks, Sci. Rep. 11 (1) (2021) 1087, https://doi.org/10.1038/s41598-020-79925-4.
[91] T. DeVries, G.W. Taylor, Leveraging Uncertainty Estimates for Predicting Segmentation Quality, 2018.
[92] G. Luo, et al., Commensal correlation network between segmentation and direct area estimation for bi-ventricle quantification, Med. Image Anal. 59 (2020), 101591, https://doi.org/10.1016/j.media.2019.101591.
[93] F.C. Ghesu, et al., Quantifying and leveraging classification uncertainty for chest radiograph assessment, Lect. Notes Comput. Sci. 11769 LNCS (2019) 676–684, https://doi.org/10.1007/978-3-030-32226-7_75.
[94] M.S. Graham, et al., Hierarchical brain parcellation with uncertainty, Lect. Notes Comput. Sci. 12443 LNCS (2020) 23–31, https://doi.org/10.1007/978-3-030-60365-6_3.
[95] Z. Liao, et al., On modelling label uncertainty in deep neural networks: automatic estimation of intra- observer variability in 2D echocardiography quality assessment, IEEE Trans. Med. Imag. 39 (6) (2020) 1868–1883, https://doi.org/10.1109/TMI.2019.2959209.
[96] X. Li, Y. Zhou, N.C. Dvornek, Y. Gu, P. Ventola, J.S. Duncan, in: A.L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M.A. Zuluaga, S.K. Zhou, D. Racoceanu, L. Joskowicz (Eds.), Efficient Shapley Explanation for Features Importance Estimation under Uncertainty BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Springer International Publishing, Cham, 2020, pp. 792–801.
[97] C. Ye, Y. Li, X. Zeng, An improved deep network for tissue microstructure estimation with uncertainty quantification, Med. Image Anal. 61 (2020), https://doi.org/10.1016/j.media.2020.101650.
[98] M. Tardy, B. Scheffer, D. Mateus, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), Uncertainty Measurements for the Reliable Classification of Mammograms BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Springer International Publishing, Cham, 2019, pp. 495–503.
[99] M.H. Jensen, D.R. Jørgensen, R. Jalaboi, M.E. Hansen, M.A. Olsen, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), Improving Uncertainty Estimation in Convolutional Neural Networks Using Inter-rater Agreement BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Springer International Publishing, Cham, 2019, pp. 540–548.
[101] Y. Gal, Z.A. Uk, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning Zoubin Ghahramani, 2016 [Online]. Available: http://yarin.co. (Accessed 7 June 2023).
[102] G.E. Hinton, D. van Camp, Keeping Neural Networks Simple by Minimizing the Description Length of the Weights, 1993, pp. 5–13, https://doi.org/10.1145/168304.168306.
[103] MethodRadford Carlo, NealTechnical, Bayesian Training of Backpropagation Networks by the Hybrid Monte, 1993.
[104] J.S. Denker, Y. LeCun, Transforming neural-net output levels to probability distributions, Adv. Neural Inf. Process. Syst. 3 (1991) 853–859.
[105] C. Leibig, V. Allken, M.S. Ayhan, P. Berens, S. Wahl, Leveraging uncertainty information from deep neural networks for disease detection, Sci. Rep. 7 (1) (2017) 1–14, https://doi.org/10.1038/s41598-017-17876-z.
[106] O. Ozdemir, B. Woodward, A.A. Berlin, Propagating Uncertainty in Multi-Stage Bayesian Convolutional Neural Networks with Application to Pulmonary Nodule Detection, Nips, 2017, pp. 1–6.
[107] A. Jungo, et al., Towards uncertainty-assisted brain tumor segmentation and survival prediction, Lect. Notes Comput. Sci. 10670 LNCS (2018) 474–485, https://doi.org/10.1007/978-3-319-75238-9_40.
[108] A. Jungo, et al., On the effect of inter-observer variability for a reliable estimation of uncertainty of medical image segmentation, Lect. Notes Comput. Sci. 11070 LNCS (2018) 682–690, https://doi.org/10.1007/978-3-030-00928-1_77.
[109] J.I. Orlando, et al., U2-net: a Bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological oct scans, Proc. Int. Symp. Biomed. Imag. (2019) 1441–1445, https://doi.org/10.1109/ISBI.2019.8759581, 2019-April, no. Isbi.
[110] J. Heo, et al., Uncertainty-aware attention for reliable interpretation and prediction, Adv. Neural Inf. Process. Syst. 2018-Decem (2018) 909–918.
[111] T.A. Adrian Tousignant, Lemaitre Paul, Doina Precup, Douglas L. Arnold, Prediction of Future Multiple Sclerosis Disease Progression Using Deep Learning Analysis of MRI Data, Proceesings of Machine Learning Research, 2019, pp. 483–492.
[112] A.G. Roy, S. Conjeti, N. Navab, C. Wachinger, Bayesian QuickNAT: model uncertainty in deep whole-brain segmentation for structure-wise quality control, Neuroimage 195 (2019) 11–22, https://doi.org/10.1016/j.neuroimage.2019.03.042.
[113] L. Herzog, E. Murina, O. Dürr, S. Wegener, B. Sick, Integrating uncertainty in deep neural networks for MRI based stroke analysis, Med. Image Anal. 65 (2020), 101790, https://doi.org/10.1016/j.media.2020.101790.
[114] C.F. Baumgartner, et al., PHiSeg: capturing uncertainty in medical image segmentation, Lect. Notes Comput. Sci. 11765 LNCS (2019) 119–127, https://doi.org/10.1007/978-3-030-32245-8_14.
[115] Ł. Rączkowski, M. Możejko, J. Zambonelli, E. Szczurek, ARA: accurate, reliable and active histopathological image classification framework with Bayesian deep learning, Sci. Rep. 9 (1) (2019) 1–12, https://doi.org/10.1038/s41598-019-50587-1.
[116] Z. Eaton-Rosen, T. Varsavsky, S. Ourselin, M.J. Cardoso, As easy as 1, 2..4? Uncertainty in counting tasks for medical imaging, Lect. Notes Comput. Sci. 11767 LNCS (2019) 356–364, https://doi.org/10.1007/978-3-030-32251-9_39.
[117] M.L. di Scandalea, C.S. Perone, M. Boudreau, J. Cohen-Adad, Deep Active Learning for Axon-Myelin Segmentation on Histology Data, 2019, pp. 1–8.
[118] R. Jena, S.P. Awate, A Bayesian neural net to segment images with uncertainty estimates and good calibration, in: A.C.S. Chung, J.C. Gee, P.A. Yushkevich, S. Bao (Eds.), Information Processing in Medical Imaging, Springer International Publishing, Cham, 2019, pp. 3–15.
[119] R.D. Soberanis-Mukul, N. Navab, S. Albarqouni, in: Uncertainty-based Graph Convolutional Networks for Organ Segmentation Refinement, 2019, pp. 1–15, 1.
[120] S. Hu, D. Worrall, S. Knegt, B. Veeling, H. Huisman, M. Welling, Supervised uncertainty quantification for segmentation with multiple annotations, Lect. Notes Comput. Sci. 11765 LNCS (2019) 137–145, https://doi.org/10.1007/978-3-030-32245-8_16.
[121] M. Combalia, F. Hueto, S. Puig, J. Malvehy, V. Vilaplana, Uncertainty estimation in deep neural networks for dermoscopic image classification, IEEE Comput. Soc. Conf. Comput. Vis. Patter Recogn. Workshops 2020-June (2020) 3211–3220, https://doi.org/10.1109/CVPRW50498.2020.00380.

[122] S. Toledo-Cortés, M.D. La Pava, O. Perd'omo, F.A. Gonz'alez, Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification, OMIA@MICCAI, 2020.
[123] M.-H. Laves, J.F. Fast, L.A. Kahrs, T. Ortmaier, Well-calibrated regression uncertainty in medical imaging with deep learning, Proc. Mach. Learn Res. 121 (2020) 393–412.
[124] X. Hu, et al., Coarse-to-Fine adversarial networks and zone-based uncertainty analysis for NK/T-Cell lymphoma segmentation in CT/PET images, IEEE J. Biomed. Health Inform. 24 (9) (2020) 2599–2608, https://doi.org/10.1109/JBHI.2020.2972694.
[125] T. Nair, D. Precup, D.L. Arnold, T. Arbel, Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation, Med. Image Anal. 59 (2020), 101557, https://doi.org/10.1016/j.media.2019.101557.
[126] Y. Kwon, J.-H. Won, B.J. Kim, M.C. Paik, Uncertainty quantification using Bayesian neural networks in classification: application to biomedical image segmentation, Comput. Stat. Data Anal. 142 (2020), 106816, https://doi.org/10.1016/j.csda.2019.106816.
[127] R. Selvan, F. Faye, J. Middleton, A. Pai, Uncertainty quantification in medical image segmentation with normalizing flows, Lect. Notes Comput. Sci. 12436 (2020) 80–90, https://doi.org/10.1007/978-3-030-59861-7_9. LNCS.
[128] P. Seebock, et al., Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT, IEEE Trans. Med. Imag. 39 (1) (2020) 87–98, https://doi.org/10.1109/TMI.2019.2919951.
[129] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano, Y. Sato, Automated muscle segmentation from clinical CT using Bayesian U-net for personalized musculoskeletal modeling, IEEE Trans. Med. Imag. 39 (4) (2020) 1030–1040, https://doi.org/10.1109/TMI.2019.2940555.
[130] Y. Xia et al., "3D Semi-supervised Learning with Uncertainty-Aware Multi-View Co-training," pp. 3646–3655..
[131] E.K. Marc Gantenbein, Ertunc Erdil, RevPHiSeg : A Memory-Efficient Neural Network for Uncertainty Quantification in Medical Image Segmentation, 2020.
[132] K. Wickstrøm, M. Kampffmeyer, R. Jenssen, Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps, Med. Image Anal. 60 (2020), https://doi.org/10.1016/j.media.2019.101619.
[133] G. Carneiro, L. Zorron Cheng Tao Pu, R. Singh, A. Burt, Deep learning uncertainty and confidence calibration for the five-class polyp classification from colonoscopy, Med. Image Anal. 62 (2020), https://doi.org/10.1016/j.media.2020.101653.
[134] B. Li, T.S. Alstrøm, On Uncertainty Estimation in Active Learning for Image Segmentation, 2020.
[135] L. Quan, Y. Li, X. Chen, N. Zhang, in: A.L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M.A. Zuluaga, S.K. Zhou, D. Racoceanu, L. Joskowicz (Eds.), An Effective Data Refinement Approach for Upper Gastrointestinal Anatomy Recognition BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Springer International Publishing, Cham, 2020, pp. 43–52.
[136] Y. Wang, et al., Double-uncertainty weighted method for semi-supervised learning, Lect. Notes Comput. Sci. 12261 LNCS (2020) 542–551, https://doi.org/10.1007/978-3-030-59710-8_53.
[137] C. Bian, et al., Uncertainty-aware domain alignment for anatomical structure segmentation, Med. Image Anal. 64 (2020), https://doi.org/10.1016/j.media.2020.101732.
[138] R. Tanno, et al., Uncertainty modelling in deep learning for safer neuroimage enhancement: demonstration in diffusion MRI, Neuroimage 225 (August 2020), 117366, https://doi.org/10.1016/j.neuroimage.2020.117366, 2021.
[139] P. Thiagarajan, P. Khairnar, S. Ghosh, Explanation and use of uncertainty quantified by Bayesian neural network classifiers for breast histopathology images, IEEE Trans. Med. Imag. 41 (4) (2022) 815–825, https://doi.org/10.1109/TMI.2021.3123300.
[140] B. Ghoshal, A. Tucker, B. Sanghera, W.L. Wong, Estimating uncertainty in deep learning for reporting confidence to clinicians in medical image segmentation and diseases detection, Comput. Intell. 37 (2) (May 2021) 701–734, https://doi.org/10.1111/COIN.12411.
[141] V. Edupuganti, M. Mardani, S. Vasanawala, J. Pauly, Uncertainty quantification in deep MRI reconstruction, IEEE Trans. Med. Imag. 40 (1) (2021) 239–250, https://doi.org/10.1109/TMI.2020.3025065.
[142] M.M.A. Valiuddin, C.G.A. Viviers, R.J.G. van Sloun, P.H.N. de With, F. van der Sommen, Improving aleatoric uncertainty quantification in multi-annotated medical image segmentation with normalizing flows, Lect. Notes Comput. Sci. 12959 (2021) 75–88, https://doi.org/10.1007/978-3-030-87735-4_8. LNCS.
[143] X. Teng, S. Pei, Y.R. Lin, StoCast: stochastic disease forecasting with progression uncertainty, IEEE J. Biomed. Health Inform. 25 (3) (2021) 850–861, https://doi.org/10.1109/JBHI.2020.3006719.
[144] X. Wang, et al., UD-MIL: uncertainty-driven deep multiple instance learning for OCT image classification, IEEE J. Biomed. Health Inform. 24 (12) (2020) 3431–3442, https://doi.org/10.1109/JBHI.2020.2983730.
[145] J. Zhang, U. Norinder, F. Svensson, Deep learning-based conformal prediction of toxicity, J. Chem. Inf. Model. 61 (6) (2021) 2648–2657, https://doi.org/10.1021/acs.jcim.1c00208.
[146] J.F. Vranken, et al., Uncertainty estimation for deep learning-based automated analysis of 12-lead electrocardiograms, Eur. Heart J. Digit. Health 2 (3) (2021) 401–415, https://doi.org/10.1093/ehjdh/ztab045.
[147] I. Sieradzki, D. Leśniak, S. Podlewska, How sure can we be about ML methods-based evaluation of compound activity: incorporation of information about prediction uncertainty using deep learning techniques, Molecules 25 (6) (2020), https://doi.org/10.3390/molecules25061452.
[148] P. Natekar, A. Kori, G. Krishnamurthi, Demystifying brain tumor segmentation networks: interpretability and uncertainty analysis, Front. Comput. Neurosci. 14 (2020) 1–12, https://doi.org/10.3389/fncom.2020.00006. February.
[149] A. Sedghi, T. Kapur, J. Luo, P. Mousavi, W.M. Wells, Probabilistic image registration via deep multi-class classification: characterizing uncertainty, Lect. Notes Comput. Sci. 11840 (2019) 12–22, https://doi.org/10.1007/978-3-030-32689-0_2. LNCS, no. March 2021.
[150] A. Norouzi, A. Emami, K. Najarian, N. Karimi, S. samavi, S.M.R. Soroushmehr, Exploiting Uncertainty of Deep Neural Networks for Improving Segmentation Accuracy in MRI Images, ICASSP, IEEE, 2019, pp. 2322–2326, https://doi.org/10.1109/ICASSP.2019.8682530.
[151] A. Filos, et al., A Systematic Comparison of Bayesian Deep Learning Robustness in Diabetic Retinopathy Tasks, NeurIPS, 2019, pp. 1–12.
[152] B. Ghoshal, A. Tucker, Estimating Uncertainty and Interpretability in Deep Learning for Coronavirus (COVID-19) Detection, 2020 [Online]. Available.
[153] J.M. Dolezal, et al., Uncertainty-informed deep learning models enable high-confidence predictions for digital histopathology, Nat. Commun. 13 (1) (2022) 6572, https://doi.org/10.1038/s41467-022-34025-x.
[154] P.K. Mensah, et al., Uncertainty estimation using variational mixture of Gaussians capsule network for health image classification, Comput. Intell. Neurosci. 2022 (2022), 4984490, https://doi.org/10.1155/2022/4984490.
[155] B. Mazoure, A. Mazoure, J. Bédard, V. Makarenkov, DUNEScan: a web server for uncertainty estimation in skin cancer detection with deep neural networks, Sci. Rep. 12 (1) (2022) 179, https://doi.org/10.1038/s41598-021-03889-2.
[156] V. Jahmunah, E.Y.K. Ng, R.-S. Tan, S.L. Oh, U.R. Acharya, Uncertainty quantification in DenseNet model using myocardial infarction ECG signals, Comput. Methods Progr. Biomed. 229 (2023), 107308, https://doi.org/10.1016/j.cmpb.2022.107308.
[157] C. Stoean, et al., Automated detection of presymptomatic conditions in spinocerebellar ataxia type 2 using Monte Carlo dropout and deep neural network techniques with electrooculogram signals, Sensors (Switzerland) 20 (11) (Jun. 2020), https://doi.org/10.3390/s20113032.
[158] F. Guo, et al., Improving cardiac MRI convolutional neural network segmentation on small training datasets and dataset shift: a continuous kernel cut approach, Med. Image Anal. 61 (Apr. 2020), https://doi.org/10.1016/j.media.2020.101636.
[159] J.H.B. da Silva, P.C. Cortez, S.K. Jagatheesaperumal, V.H.C. de Albuquerque, ECG measurement uncertainty based on Monte Carlo approach: an effective analysis for a successful cardiac health monitoring system, Bioengineering 10 (1) (Jan. 2023), https://doi.org/10.3390/bioengineering10010115.
[160] T. Nasir, M.K. Malik, SACDNet: towards Early Type 2 Diabetes Prediction with Uncertainty for Electronic Health Records, Jan. 2023 [Online]. Available: http://arxiv.org/abs/2301.04844.
[161] M. Abdar, et al., UncertaintyFuseNet: robust uncertainty-aware hierarchical feature fusion model with Ensemble Monte Carlo Dropout for COVID-19 detection, Inf. Fusion 90 (Feb. 2023) 364–381, https://doi.org/10.1016/j.inffus.2022.09.023.
[162] S. MacDonald, et al., Generalising uncertainty improves accuracy and safety of deep learning analytics applied to oncology, Sci. Rep. 13 (1) (May 2023) 7395, https://doi.org/10.1038/s41598-023-31126-5.
[163] M.U. Farooq, Z. Ullah, J. Gwak, Residual attention based uncertainty-guided mean teacher model for semi-supervised breast masses segmentation in 2D ultrasonography, Comput. Med. Imag. Graph. 104 (Mar. 2023), 102173, https://doi.org/10.1016/J.COMPMEDIMAG.2022.102173.
[164] M. Abdar, et al., Hercules: deep hierarchical attentive multilevel fusion model with uncertainty quantification for medical image classification, IEEE Trans. Ind. Inf. 19 (1) (Jan. 2023) 274–285, https://doi.org/10.1109/TII.2022.3168887.
[165] A. Zakeri, et al., DragNet: learning-based deformable registration for realistic cardiac MR sequence generation from a single frame, Med. Image Anal. 83 (Jan. 2023), https://doi.org/10.1016/j.media.2022.102678.
[166] M. Abdar, et al., Binarized multi-gate mixture of Bayesian experts for cardiac syndrome X diagnosis: a clinician-in-the-loop scenario with a belief-uncertainty fusion paradigm, Inf. Fusion 97 (Sep. 2023), https://doi.org/10.1016/j.inffus.2023.101813.
[167] R. Tanno, et al., Uncertainty Quantification in Deep Learning for Safer Neuroimage Enhancement, Jul. 2019 [Online]. Available: http://arxiv.org/abs/1907.13418.
[168] Z. Wang, et al., Bayesian Logistic Shape Model Inference : Application to Cochlear Image Segmentation, 2021.
[169] C. Corrado, et al., Quantifying atrial anatomy uncertainty from clinical data and its impact on electro-physiology simulation predictions, Med. Image Anal. 61 (Apr. 2020), https://doi.org/10.1016/j.media.2019.101626.
[170] J. Dhamala, et al., Quantifying the Uncertainty in Model Parameters Using Gaussian Process-Based Markov Chain Monte Carlo in Cardiac Electrophysiology, 2018.
[171] J. Chen, E.C. Frey, Y. He, W.P. Segars, Y. Li, Y. Du, TransMorph: Transformer for Unsupervised Medical Image Registration, Nov. 2021, https://doi.org/10.1016/j.media.2022.102615.
[172] A.A. Abdullah, M.M. Hassan, Y.T. Mustafa, Uncertainty quantification for MLP-mixer using Bayesian deep learning, Appl. Sci. 13 (7) (2023), https://doi.org/10.3390/app13074547. Apr.
[173] J. M. Dolezal et al., "Slideflow: Deep Learning for Digital Histopathology with Real-Time Whole-Slide Visualization.".
[174] K. Chitta, J.M. Alvarez, A. Lesnikowski, Large-Scale Visual Active Learning with Deep Probabilistic Ensembles, 2018.

[175] L. Smith, Y. Gal, Understanding measures of uncertainty for adversarial example detection, in: 34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018 vol. 2, 2018, pp. 560–569.
[176] A. Jungo, M. Reyes, Assessing reliability and challenges of uncertainty estimations for medical image segmentation, Lect. Notes Comput. Sci. 11765 LNCS (2019) 48–56, https://doi.org/10.1007/978-3-030-32245-8_6.
[177] P. McClure, et al., Distributed weight consolidation: a brain segmentation case study, Adv. Neural Inf. Process. Syst. 2018-Decem (2018) 4093–4103, no. NeurIPS 2018.
[178] Q. Wu, H. Li, L. Li, Z. Yu, Quantifying Intrinsic Uncertainty in Classification via Deep Dirichlet Mixture Networks, 2019.
[179] J. Linmans, J. Van Der Laak, G. Litjens, Efficient out-of-distribution detection in digital pathology using multi-head convolutional neural networks, Proc. Mach. Learn Res. 121 (2020) 465–478.
[180] G. Liang, Y. Zhang, N. Jacobs, Neural network calibration for medical imaging classification using DCA regularization, ICML Workshop Uncertain, in: Robustness Deep Learn. presented in Workshop Uncertainty and Robustness in Deep Learning Workshop (ICML 2020), 2020.
[181] K. Hoebel, et al., An exploration of uncertainty information for segmentation quality assessment, Proc. SPIE (Mar. 2020), https://doi.org/10.1117/12.2548722.
[182] A. Mehrtash, W.M. Wells, C.M. Tempany, P. Abolmaesumi, T. Kapur, Confidence calibration and predictive uncertainty estimation for deep medical image segmentation, IEEE Trans. Med. Imag. 39 (12) (2020) 3868–3878, https://doi.org/10.1109/TMI.2020.3006437.
[183] L. Dahal, A. Kafle, B. Khanal, Echocardiography SegmentationUncertainty Estimation in Deep 2D, May, 2020.
[184] E. Chiou, F. Giganti, S. Punwani, I. Kokkinos, E. Panagiotaki, Harnessing uncertainty in domain adaptation for mri prostate lesion segmentation, Lect. Notes Comput. Sci. 12261 LNCS (2020) 510–520, https://doi.org/10.1007/978-3-030-59710-8_50.
[185] X. Cao, H. Chen, Y. Li, Y. Peng, S. Wang, L. Cheng, Uncertainty aware temporal-ensembling model for semi-supervised ABUS mass segmentation, IEEE Trans. Med. Imag. 40 (1) (2021) 431–443, https://doi.org/10.1109/TMI.2020.3029161.
[186] Y. Qin, Z. Liu, C. Liu, Y. Li, X. Zeng, C. Ye, Super-Resolved q-Space deep learning with uncertainty quantification, Med. Image Anal. 67 (2021), 101885, https://doi.org/10.1016/j.media.2020.101885.
[187] R.K. Singh, R. Gorantla, S.G.R. Allada, P. Narra, SkiNet: a deep learning framework for skin lesion diagnosis with uncertainty estimation and explainability, PLoS One 17 (10) (Oct. 2022), e0276836.
[188] F. Guo, M. Ng, G. Kuling, G. Wright, Cardiac MRI segmentation with sparse annotations: ensembling deep learning uncertainty and shape priors, Med. Image Anal. 81 (Oct) (2022), https://doi.org/10.1016/j.media.2022.102532.
[189] T. Buddenkotte, et al., Calibrating ensembles for scalable uncertainty quantification in deep learning-based medical image segmentation, Comput. Biol. Med. 163 (Sep. 2023), https://doi.org/10.1016/j.compbiomed.2023.107096.
[190] D. Molchanov, A. Lyzhov, Y. Molchanova, A. Ashukha, D. Vetrov, Greedy policy search : a simple baseline for learnable test-time augmentation 2 (7) (2020).
[191] G. Wang, et al., Interactive medical image segmentation using deep learning with image-specific fine tuning, IEEE Trans. Med. Imag. 37 (7) (2018) 1562–1573, https://doi.org/10.1109/TMI.2018.2791721.
[192] M.S. Ayhan, P. Berens, Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks, Med. Imag. Deep Learn. (MIDL), no. Midl (2018) 1–9.
[193] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks, Neurocomputing 338 (2019) 34–45, https://doi.org/10.1016/j.neucom.2019.01.103.
[194] Z. Zhang, A. Romero, M.J. Muckley, P. Vincent, L. Yang, M. Drozdzal, Reducing uncertainty in undersampled MRI reconstruction with active acquisition, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. (2019-June) 2049–2053, https://doi.org/10.1109/CVPR.2019.00215, 2019.
[195] C. Athanasiadis, E. Hortal, S. Asteriadis, Audio–visual domain adaptation using conditional semi-supervised Generative Adversarial Networks, Neurocomputing 397 (xxxx) (2020) 331–344, https://doi.org/10.1016/j.neucom.2019.09.106.
[196] M.S. Ayhan, L. Kühlewein, G. Aliyeva, W. Inhoffen, F. Ziemssen, P. Berens, Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection, Med. Image Anal. 64 (2020), 101724, https://doi.org/10.1016/j.media.2020.101724.
[197] T. Araújo, et al., DR|GRADUATE: uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images, Med. Image Anal. 63 (2020), 101715, https://doi.org/10.1016/j.media.2020.101715.
[198] M. Abdar, et al., Uncertainty quantification in skin cancer classification using three-way decision-based Bayesian deep learning, Comput. Biol. Med. 135 (2021), 104418, https://doi.org/10.1016/j.compbiomed.2021.104418.
[199] G. Scalia, C.A. Grambow, B. Pernici, Y.-P. Li, W.H. Green, Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction, J. Chem. Inf. Model. 60 (6) (Jun. 2020) 2697–2717, https://doi.org/10.1021/acs.jcim.9b00975.
[200] S. Dong, Q. Yang, Y. Fu, M. Tian, C. Zhuo, RCoNet: deformable mutual information maximization and high-order uncertainty-aware learning for robust COVID-19 detection, IEEE Transact. Neural Networks Learn. Syst. 32 (8) (Aug. 2021) 3401–3411, https://doi.org/10.1109/TNNLS.2021.3086570.
[201] I. Cortés-Ciriano, A. Bender, Deep confidence: a computationally efficient framework for calculating reliable prediction errors for deep neural networks, J. Chem. Inf. Model. 59 (3) (2019) 1269–1281, https://doi.org/10.1021/acs.jcim.8b00542.
[202] I. Cortés-Ciriano, A. Bender, Reliable prediction errors for deep neural networks using test-time dropout, J. Chem. Inf. Model. 59 (7) (Jul. 2019) 3330–3339, https://doi.org/10.1021/acs.jcim.9b00297.
[203] F. Hamedani-KarAzmoudehFar, R. Tavakkoli-Moghaddam, A.R. Tajally, S.S. Aria, Breast cancer classification by a new approach to assessing deep neural network-based uncertainty quantification methods, Biomed. Signal Process Control 79 (Jan. 2023), https://doi.org/10.1016/j.bspc.2022.104057.
[204] B.L. Graham, et al., Standardization of spirometry 2019 update an official American Thoracic Society and European Respiratory Society technical statement, Am. J. Respir. Crit. Care Med. 200 (8) (Oct. 15, 2019) E70–E88, https://doi.org/10.1164/rccm.201908-1590ST. American Thoracic Society.
[205] S. Flügge, S. Zimmer, U. Petersohn, Knowledge Representation and Diagnostic Inference Using Bayesian Networks in the Medical Discourse, 2019.
[206] H. Li, H. Luo, Uncertainty quantification in medical image segmentation, in: 2020 IEEE 6th International Conference on Computer and Communications, ICCC), 2020, pp. 1936–1940, https://doi.org/10.1109/ICCC51575.2020.9345043.
[207] A.S. Lundervold, A. Lundervold, An overview of deep learning in medical imaging focusing on MRI, Z. Med. Phys. 29 (2) (2019) 102–127, https://doi.org/10.1016/j.zemedi.2018.11.002.
[208] A. Kamal, et al., Recent advances and challenges in uncertainty visualization: a survey, J. Vis. 24 (5) (2021) 861–890, https://doi.org/10.1007/s12650-021-00755-1.
[209] D. Milanés-Hermosilla, et al., Monte Carlo dropout for uncertainty estimation and motor imagery classification, Sensors 21 (21) (2021), https://doi.org/10.3390/s21217241.
[210] E. Begoli, T. Bhattacharya, D. Kusnezov, The need for uncertainty quantification in machine-assisted medical decision making, Nat. Mach. Intell. 1 (1) (Jan. 01, 2019) 20–23, https://doi.org/10.1038/s42256-018-0004-1. Nature Research.
[211] M.W. Nadeem, H.G. Goh, V. Ponnusamy, I. Andonovic, M.A. Khan, M. Hussain, A fusion-based machine learning approach for the prediction of the onset of diabetes, Healthcare (Switzerland) 9 (10) (Oct. 2021), https://doi.org/10.3390/healthcare9101393.
[212] B. Ihnaini, et al., A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning, 2021, https://doi.org/10.1155/2021/4243700.
[213] Y. Yao, "LNAI 7413 - an Outline of a Theory of Three-Way Decisions.".
[214] M.G. Uddin, S. Nash, A. Rahman, A.I. Olbert, A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches, Water Res. 229 (Feb. 2023), https://doi.org/10.1016/j.watres.2022.119422.
[215] N. Molchanova, et al., Novel Structural-Scale Uncertainty Measures and Error Retention Curves: Application to Multiple Sclerosis, Nov. 2022, https://doi.org/10.48550/ARXIV.2211.04825.
[216] T.N. Kipf, M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, Sep. 2016 [Online]. Available: http://arxiv.org/abs/1609.02907.
[217] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering.".[Online]. Available: https://github.com/mdeff/cnn_graph.
[218] G.C. Al, A machine learning approach for the differential diagnosis of alzheimer and vascular dementia fed by MRI selected features, Front. Neuroinf. 14 (Jun) (2020), https://doi.org/10.3389/fninf.2020.00025.
[219] S.K. Sood, S. Kaur, K.K. Chahal, An intelligent framework for monitoring dengue fever risk using LDA-ANFIS, J. Ambient Intell. Smart Environ. 12 (1) (2020) 5–20, https://doi.org/10.3233/AIS-200547.
[220] M.Y.T. al, A new open-source GPU-based microscopic Monte Carlo simulation tool for the calculations of DNA damages caused by ionizing radiation — Part I: core algorithm and validation, Med. Phys. 47 (4) (Apr. 2020) 1958–1970, https://doi.org/10.1002/MP.14037.
[221] M.V.S. al, Projected impact of a reduction in sugar-sweetened beverage consumption on diabetes and cardiovascular disease in Argentina: a modeling study, PLoS Med. 17 (7) (Jul. 2020), https://doi.org/10.1371/JOURNAL.PMED.1003224.
[222] A.G. al, Mixed-effects models for health care longitudinal data with an informative visiting process: a Monte Carlo simulation study, Stat. Neerl. 74 (1) (Feb. 2020) 5–23, https://doi.org/10.1111/STAN.12188.
[223] X.L. Sun, H. Wang, X.K. Li, G.H. Cao, Y. Kuang, X.C. Zhang, Monte Carlo computer simulation of a camera system for proton beam range verification in cancer treatment, Future Generat. Comput. Syst. 102 (Jan. 2020) 978–991, https://doi.org/10.1016/J.FUTURE.2019.09.011.
[224] B.P. Magnusson, H. Schmidli, N. Rouyrre, D.O. Scharfstein, Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by postrandomization event occurrence, Stat. Med. 38 (23) (Oct. 2019) 4761–4771, https://doi.org/10.1002/sim.8333.
[225] K.L. al, Big medical data decision-making intelligent system exploiting fuzzy inference logic for prostate cancer in developing countries, IEEE Access 7 (2019) 2348–2363, https://doi.org/10.1109/ACCESS.2018.2886198.
[226] S. Razi, M.R.K. Mollaei, J. Ghasemi, A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory, Inf. Sci. 484 (May 2019) 14–26, https://doi.org/10.1016/J.INS.2019.01.053.